

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04018
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,580
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,583
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,581
; FILING DATE: 01-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: CCOR-0232
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-04018-58

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
Db 5 QPTPEPS 11

RESULT 23
PCT-US95-04018-59
; Sequence 59, Application PC/TUS9504018
; GENERAL INFORMATION:
; APPLICANT: Heavenr, George A.
; APPLICANT: Kruszynski, Marian
; APPLICANT: Mervic, Miljenko
; APPLICANT: Weber, Robert W.
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04018
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,580
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/221,583
; FILING DATE: 01-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/221,581
; FILING DATE: 01-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: CCOR-0232
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-04018-59

Query Match 0.9%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358
Db 2 QPTPEPS 8

RESULT 24
US-09-020-116-4
; Sequence 4, Application US/09020116
; Patent No. 6084063
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS
; TITLE OF INVENTION: OF ALLERGIC REACTIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: SUITE 1200, 127 PEACHTREE STREET
; CITY: ATLANTA
; STATE: GA
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLER, MARY L
; REGISTRATION NUMBER: 39,303
; REFERENCE/DOCKET NUMBER: 14014.0285
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-221-583-58

Query Match 0.9%; Score 7; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358  
Db 5 QPTPEPS 11

RESULT 20  
US-08-221-583-59  
Sequence 59, Application US/08221583  
Patent No. 5486595  
GENERAL INFORMATION:  
APPLICANT: Heavner, George A.  
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595ris  
STREET: One Liberty Place 46th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcmMod.  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/221,583  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: DeLuca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: CCOR-0185  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-221-583-59

Query Match 0.9%; Score 7; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358  
Db 2 QPTPEPS 8

RESULT 21  
PCT-US95-04018-57  
Sequence 57, Application PC/TUS9504018  
GENERAL INFORMATION:  
APPLICANT: Heavner, George A.

APPLICANT: Kruszynski, Marian  
APPLICANT: Mervic, Miljenko  
APPLICANT: Weber, Robert W.  
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
STREET: One Liberty Place 46th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04018  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/221,580  
FILING DATE: 01-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/221,583  
FILING DATE: 01-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/221,581  
FILING DATE: 01-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: DeLuca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: CCOR-0232  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-04018-57

Query Match 0.9%; Score 7; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 QPTPEPS 358  
Db 8 QPTPEPS 14

RESULT 22  
PCT-US95-04018-58  
Sequence 58, Application PC/TUS9504018  
GENERAL INFORMATION:  
APPLICANT: Heavner, George A.  
APPLICANT: Kruszynski, Marian  
APPLICANT: Mervic, Miljenko  
APPLICANT: Weber, Robert W.  
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
STREET: One Liberty Place 46th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19403

; LENGTH: 490  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5116

Query Match 1.0%; Score 8; DB 4; Length 490;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 NDGAVALA 157  
Db 72 NDGAVALA 79

RESULT 17  
US-08-999-774A-10  
; Sequence 10, Application US/08999774A  
; Patent No. 6274312  
; GENERAL INFORMATION:  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Seghezzi, Wolfgang  
; APPLICANT: Shanahan, Frances  
; APPLICANT: Lees, Emma M.  
; APPLICANT: McClanahan, Terrill K.  
; TITLE OF INVENTION: Intracellular Regulatory Molecules;  
; TITLE OF INVENTION: Related Reagents  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,774A  
; FILING DATE: 10-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/032,818  
; FILING DATE: 11-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0646  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)852-9196  
; TELEFAX: (650)496-1200  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 503 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-999-774A-10

Query Match 1.0%; Score 8; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 357 PSPGPQPA 364  
Db 244 PSPGPQPA 251

RESULT 18  
US-08-221-583-57  
; Sequence 57, Application US/08221583

; Patent No. 5486595  
; GENERAL INFORMATION:  
; APPLICANT: Heavner, George A.  
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595tris  
; STREET: One Liberty Place 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcMod.  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/221,583  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: CCOR-0185  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-221-583-57

Query Match 0.9%; Score 7; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 QPTPEPS 358  
Db 8 QPTPEPS 14

RESULT 19  
US-08-221-583-58  
; Sequence 58, Application US/08221583  
; Patent No. 5486595  
; GENERAL INFORMATION:  
; APPLICANT: Heavner, George A.  
; TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595tris  
; STREET: One Liberty Place 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcMod.  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/221,583  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: CCOR-0185

Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TMTSNNST 257  
|||||  
Db 178 TMTSNNST 186

## RESULT 14

US-08-569-166-34  
; Sequence 34, Application US/08569166  
; Patent No. 5830722  
; GENERAL INFORMATION:  
; APPLICANT: NICOLAS, LUC  
; APPLICANT: CHARLES, JEAN-FRANCOIS  
; APPLICANT: DELECLUSE, ARMELE  
; APPLICANT: BARLOY, FREDERIQUE  
; TITLE OF INVENTION: CLOSTRIDIUM BIFERMENTANS DNA FRAGMENT  
; TITLE OF INVENTION: BEARING GENES CODING FOR PROTEINS LINKED TO AN  
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/569,166  
; FILING DATE: 05-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR94/00768  
; FILING DATE: 24-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 93/07795  
; FILING DATE: 25-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 660-106-0 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 110 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-569-166-34

Query Match 1.0%; Score 8; DB 2; Length 110;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 TMTSNNST 257  
|||||  
Db 67 TMTSNNST 74

## RESULT 15

US-09-107-532A-5477  
; Sequence 5477, Application US/09107532A  
; Patent No. 6583275

; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5477:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 332 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...332  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5477:  
; US-09-107-532A-5477

Query Match 1.0%; Score 8; DB 4; Length 332;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 EEINRQKQ 137  
|||||  
Db 137 EEINRQKQ 144

## RESULT 16

US-09-134-001C-5116  
; Sequence 5116, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5116



```
Query Match          4.0%; Score 32; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 5.2e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDGYVTSHGDHYHYNGKVPYDA 77
    |||||||
Db 67 AEQIVIKITDGYVTSHGDHYHYNGKVPYDA 98

RESULT 11
US-09-328-352-6143
; Sequence 6143, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6143
; LENGTH: 795
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6143

Query Match          1.1%; Score 9; DB 4; Length 795;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 SASLAAAE 209
    |||||||
Db 428 SASLAAAE 436

RESULT 12
US-08-188-582-2
; Sequence 2, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO

Query Match          1.1%; Score 9; DB 1; Length 921;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 249 TINTSNNST 257
    |||||||
Db 178 TINTSNNST 186

RESULT 13
US-08-646-715-2
; Sequence 2, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-582-2

Query Match          1.1%; Score 9; DB 1; Length 921;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 249 TINTSNNST 257
    |||||||
Db 178 TINTSNNST 186

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-2

Query Match          1.1%; Score 9; DB 1; Length 921;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 249 TINTSNNST 257
    |||||||
Db 178 TINTSNNST 186

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-2
```

```
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 763 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

Query Match          6.4%; Score 51; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.2e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 QGRYTTDDGVIFNASDIIEDTGDAYIVPHGDHYHYIPKNLSASELAARAA 210
DB 159 QGRYTTDDGVIFNASDIIEDTGDAYIVPHGDHYHYIPKNLSASELAARAA 209

RESULT 8
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
;   APPLICANT: Choi et. al.
;   TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;   NUMBER OF SEQUENCES: 452
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Human Genome Sciences, Inc.
;     STREET: 9410 Key West Avenue
;     CITY: Rockville
;     STATE: Maryland
;     COUNTRY: USA
;     ZIP: 20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;     COMPUTER: HP Vectra 486/33
;     OPERATING SYSTEM: MSDOS version 6.2
;     SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/961,083
;     FILING DATE: 30-Oct-1997
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/961,083
;     FILING DATE: OCT-30-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Michelle S. Marks
;       REGISTRATION NUMBER: 41,971
;       REFERENCE/DOCKET NUMBER: PB340P3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (301) 309-8504
;       TELEFAX: (301) 309-8512
;     INFORMATION FOR SEQ ID NO: 182:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 447 amino acids
;         TYPE: amino acid
;         STRANDEDNESS: single
;         TOPOLOGY: linear
;         MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match          4.0%; Score 32; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 4.8e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEQIVIKITDQGYVTSBGHDHYHYNGKVPYDA 77
DB 43 AEQIVIKITDQGYVTSBGHDHYHYNGKVPYDA 74

RESULT 10
US-09-468-656A-6
; Sequence 6, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
;   APPLICANT: Johnson, Leslie S.
;   APPLICANT: Adamou, John E.
;   TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
;   TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
;   TITLE OF INVENTION: Motifs
;   FILE REFERENCE: 469201-444
;   CURRENT APPLICATION NUMBER: US/09/468,656A
;   CURRENT FILING DATE: 1999-12-02
;   PRIOR APPLICATION NUMBER: 60/113,048
;   PRIOR FILING DATE: 1998-12-21
;   NUMBER OF SEQ ID NOS: 14
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 6
;   LENGTH: 484
;   TYPE: PRI
;   ORGANISM: Streptococcus pneumoniae
US-09-468-656A-6
```



1 SYELGLYQARTVKENNRVSVIDGKQATQKTENLTPEDEVSKREGINAEQIVIKITDQGYVT 60  
1 SYELGLYQARTVKENNRVSVIDGKQATQKTENLTPEDEVSKREGINAEQIVIKITDQGYVT 60  
61 SHGDHYHYNGKVPYDAIISEELLMDPNKYLKDEDIVNEVKGYYIVKVDGKYVYVLKDA 120  
61 SHGDHYHYNGKVPYDAIISEELLMDPNKYLKDEDIVNEVKGYYIVKVDGKYVYVLKDA 120  
121 AHADNVRTKEEINRQKOEHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180  
121 AHADNVRTKEEINRQKOEHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180  
181 GDYIIVPHGDHYHYIPKVELSASELAFLSGRGNLSRSTYRRQNSDNTSRTNWVPS 240  
181 GDYIIVPHGDHYHYIPKVELSASELAFLSGRGNLSRSTYRRQNSDNTSRTNWVPS 240  
241 VSNPGTTNTNTSNTNSQASQSDNIDSLKQYKLPQRSQHVESDGLVDFDPAQITSTRT 300  
241 VSNPGTTNTNTSNTNSQASQSDNIDSLKQYKLPQRSQHVESDGLVDFDPAQITSTRT 300  
301 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPDSRPPQSPQPTPEPSPG 360  
301 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPDSRPPQSPQPTPEPSPG 360  
361 PQAPNLKIDNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLPSFVTKNLESKLSQES 420  
361 PQAPNLKIDNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLPSFVTKNLESKLSQES 420  
421 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDPQALDKLLERLNDESTN 480  
421 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDPQALDKLLERLNDESTN 480  
481 KEKLVDDLLAFAPITHPERLKGPNQSIETDEVRITAOADKYTTSDGYIFDEHDIISD 540  
481 KEKLVDDLLAFAPITHPERLKGPNQSIETDEVRITAOADKYTTSDGYIFDEHDIISD 540  
541 EGDYVTPHMGHSHWICKSLSDEKVAQAAYTKKEGILPSPDADYKANPTGDSAAIY 600  
541 EGDYVTPHMGHSHWICKSLSDEKVAQAAYTKKEGILPSPDADYKANPTGDSAAIY 600  
601 NRVKGEKRIPLVRLPYMVEHTVEVNGNLIIPHKHVHNIKFAWFDHHTYKAPNGYTTLED 660  
601 NRVKGEKRIPLVRLPYMVEHTVEVNGNLIIPHKHVHNIKFAWFDHHTYKAPNGYTTLED 660  
661 LFATIKYVVEHPDERPHSDNGWGNASHVLGKKDSEDPNKFNKFADEPVEETPAEPEVP 720  
661 LFATIKYVVEHPDERPHSDNGWGNASHVLGKKDSEDPNKFNKFADEPVEETPAEPEVP 720  
721 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780  
721 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780  
781 LLKGSNPSSVSKEKIN 796  
781 LLKGSNPSSVSKEKIN 796

## RESULT 3

US-09-468-656A-8  
; Sequence 8, Application US/09468656A  
; Patent No. 6582706  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-444  
; CURRENT APPLICATION NUMBER: US/09/468, 656A  
; CURRENT FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: 60/113, 048  
; PRIOR FILING DATE: 1998-12-21

1 NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 819  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-468-656A-8

Query Match 87.3%; Score 695; DB 4; Length 819;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSVIDGKQATQKTENLTPEDEVSKREGINAEQIVIKITDQGYVT 60  
Db 21 SYELGLYQARTVKENNRVSVIDGKQATQKTENLTPEDEVSKREGINAEQIVIKITDQGYVT 80  
QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNKYLKDEDIVNEVKGYYIVKVDGKYVYVLKDA 120  
Db 81 SHGDHYHYNGKVPYDAIISEELLMDPNKYLKDEDIVNEVKGYYIVKVDGKYVYVLKDA 140  
QY 121 AHADNVRTKEEINRQKOEHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180  
Db 141 AHADNVRTKEEINRQKOEHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 200  
QY 181 GDYIIVPHGDHYHYIPKVELSASELAFLSGRGNLSRSTYRRQNSDNTSRTNWVPS 240  
Db 201 GDYIIVPHGDHYHYIPKVELSASELAFLSGRGNLSRSTYRRQNSDNTSRTNWVPS 260  
QY 241 VSNPGTTNTNTSNTNSQASQSDNIDSLKQYKLPQRSQHVESDGLVDFDPAQITSTRT 300  
Db 261 VSNPGTTNTNTSNTNSQASQSDNIDSLKQYKLPQRSQHVESDGLVDFDPAQITSTRT 320  
QY 301 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPDSRPPQSPQPTPEPSPG 360  
Db 321 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPDSRPPQSPQPTPEPSPG 380  
QY 361 PQAPNLKIDNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLPSFVTKNLESKLSQES 420  
Db 381 PQAPNLKIDNSLSVQLVRKVGEGYVFEKIGISRYVFAKDLPSFVTKNLESKLSQES 440  
QY 421 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDPQALDKLLERLNDESTN 480  
Db 441 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDPQALDKLLERLNDESTN 500  
QY 481 KEKLVDDLLAFAPITHPERLKGPNQSIETDEVRITAOADKYTTSDGYIFDEHDIISD 540  
Db 501 KEKLVDDLLAFAPITHPERLKGPNQSIETDEVRITAOADKYTTSDGYIFDEHDIISD 560  
QY 541 EGDYVTPHMGHSHWICKSLSDEKVAQAAYTKKEGILPSPDADYKANPTGDSAAIY 600  
Db 561 EGDYVTPHMGHSHWICKSLSDEKVAQAAYTKKEGILPSPDADYKANPTGDSAAIY 620  
QY 601 NRVKGEKRIPLVRLPYMVEHTVEVNGNLIIPHKHVHNIKFAWFDHHTYKAPNGYTTLED 660  
Db 621 NRVKGEKRIPLVRLPYMVEHTVEVNGNLIIPHKHVHNIKFAWFDHHTYKAPNGYTTLED 680  
QY 661 LFATIKYVVEHPDERPHSDNGWGNASHVLGKKDSEDPNKFNKFADEPVEETPAEPEVP 720  
Db 681 LFATIKYVVEHPDERPHSDNGWGNASHVLGKKDSEDPNKFNKFADEPVEETPAEPEVP 740  
QY 721 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780  
Db 741 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 800  
QY 781 LLKGSNPSSVSKEKIN 796  
Db 801 LLKGSNPSSVSKEKIN 816

## RESULT 4

US-09-468-656A-10  
; Sequence 10, Application US/09468656A  
; Patent No. 6582706

```

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-56

Query Match 99.9%; Score 795; DB 3; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYELGLQARTVKNNRVSIDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 1 SYELGLQARTVKNNRVSIDGQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Qy 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDIVNEVKGKGVYIKVDGKYVYVYLKDA 120
Db 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDIVNEVKGKGVYIKVDGKYVYVYLKDA 120
Qy 121 AHADNVRTKEINRQKQEHSHQREGGTPRNDGAVALARPSQGRYTTDDGYIFNASDIIEDT 180
Db 121 AHADNVRTKEINRQKQEHSHQREGGTPRNDGAVALARPSQGRYTTDDGYIFNASDIIEDT 180
Qy 181 GDVIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db 181 GDVIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Qy 241 VSNPQTNTNTSNNSNTNSQASQNSNDISLLKQLYKLPLSRHVESDGLVDPDPAQITSR 300
Db 241 VSNPQTNTNTSNNSNTNSQASQNSNDISLLKQLYKLPLSRHVESDGLVDPDPAQITSR 300
Qy 301 ARGVAVPHGDHYHYIPYSQMSLEBERIARIIPLYRSNHHVDPDRPEOPSQPTPEPSPG 360
Db 301 ARGVAVPHGDHYHYIPYSQMSLEBERIARIIPLYRSNHHVDPDRPEOPSQPTPEPSPG 360
Qy 361 PQAPNLKIDNSNLSVLSQVRKVGSGYVFEKSGISRYVFAKDLPSSETVKNLESKLSKQES 420
Db 361 PQAPNLKIDNSNLSVLSQVRKVGSGYVFEKSGISRYVFAKDLPSSETVKNLESKLSKQES 420
Qy 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480

```

```

Qy 481 KEKLVDDLLAFAPITHPBELGKPNSEIYETDEVRIQAQADKYTTSDGYIFDEHDIISD 540
Db 481 KEKLVDDLLAFAPITHPBELGKPNSEIYETDEVRIQAQADKYTTSDGYIFDEHDIISD 540
Qy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTKERGIPLPPSPDADVKANPTGDSAAAY 600
Db 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTKERGIPLPPSPDADVKANPTGDSAAAY 600
Qy 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPKQDHYHNIKFAWFDHHTYKAPNGYTL 660
Db 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPKQDHYHNIKFAWFDHHTYKAPNGYTL 660
Qy 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKKDSEDPNKNFKADDEPVEETPAEPEVP 720
Db 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKKDSEDPNKNFKADDEPVEETPAEPEVP 720
Qy 721 QVETEKVEAQLKEAEVLLAKVTSSILKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780
Db 721 QVETEKVEAQLKEAEVLLAKVTSSILKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780
Qy 781 LLKGSNPSSVSKEKIN 796
Db 781 LLKGSNPSSVSKEKIN 796

RESULT 2
US-09-536-784-56
; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-536-784-56

```

Query Match 99.9%; Score 795; DB 4; Length 796;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:31:07 ; Search time 24 Seconds  
(without alignments)  
1712.261 Million cell updates/sec

Title: US-09-765-271-56  
Perfect score: 796  
Sequence: 1 SYELGLYQARTVKENRVSY.....KLLALLKGSNPSSVSKEKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/prodata/2/iaa/6C\_COMB.pep.\*
- 6: /cgn2\_6/prodata/2/iaa/6D\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	3	US-08-961-083-56
2	795	99.9	796	4	US-09-536-784-56
3	695	87.3	819	4	US-09-468-656A-8
4	60	7.5	819	4	US-09-468-656A-10
5	60	7.5	838	4	US-09-468-656A-4
6	51	6.4	763	3	US-08-961-083-66
7	51	6.4	763	4	US-09-536-784-66
8	32	4.0	447	3	US-08-961-083-182
9	32	4.0	447	4	US-09-536-784-182
10	32	4.0	484	4	US-09-468-656A-6
11	9	1.1	795	4	US-09-328-352-6143
12	9	1.1	921	1	US-08-188-582-2
13	9	1.1	921	1	US-08-546-715-2
14	8	1.0	110	2	US-08-569-166-34
15	8	1.0	332	4	US-09-107-532A-5477
16	8	1.0	490	4	US-09-134-001C-5116
17	8	1.0	503	3	US-08-999-774A-10
18	7	0.9	15	1	US-08-221-583-57
19	7	0.9	15	1	US-08-221-583-58
20	7	0.9	15	1	US-08-221-583-59
21	7	0.9	15	5	PCT-US95-04018-57
22	7	0.9	15	5	PCT-US95-04018-58
23	7	0.9	15	5	PCT-US95-04018-59
24	7	0.9	27	3	US-09-020-116-4
25	7	0.9	27	4	US-09-608-902-4
26	7	0.9	45	3	US-09-020-116-2
27	7	0.9	45	4	US-09-608-902-2

## ALIGNMENTS

RESULT 1  
US-08-961-083-56  
; Sequence 56, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:

28	7	0.9	50	3	US-09-171-646-1	Sequence 1, Appli
29	7	0.9	66	3	US-09-020-116-1	Sequence 1, Appli
30	7	0.9	66	4	US-09-608-902-1	Sequence 1, Appli
31	7	0.9	76	4	US-09-328-352-6019	Sequence 6019, Ap
32	7	0.9	117	4	US-09-134-001C-4585	Sequence 4585, Ap
33	7	0.9	141	4	US-09-543-681A-5683	Sequence 5683, Ap
34	7	0.9	149	4	US-09-252-991A-28698	Sequence 28698, A
35	7	0.9	158	2	US-08-917-456-2	Sequence 2, Appli
36	7	0.9	158	3	US-09-229-804-2	Sequence 2, Appli
37	7	0.9	163	4	US-09-252-991A-20622	Sequence 20622, A
38	7	0.9	191	4	US-09-252-991A-30217	Sequence 30217, A
39	7	0.9	198	4	US-09-613-303-35	Sequence 35, Appl
40	7	0.9	198	4	US-10-267-311-35	Sequence 35, Appl
41	7	0.9	201	2	US-08-916-901-3	Sequence 3, Appli
42	7	0.9	201	2	US-08-916-901-8	Sequence 3, Appli
43	7	0.9	201	4	US-09-154-602-3	Sequence 8, Appli
44	7	0.9	201	4	US-09-154-602-8	Sequence 8, Appli
45	7	0.9	226	3	US-08-908-332-5	Sequence 5, Appli
46	7	0.9	235	4	US-09-326-394-4	Sequence 4, Appli
47	7	0.9	235	4	US-09-580-235-2	Sequence 2, Appli
48	7	0.9	235	4	US-09-580-235-4	Sequence 4, Appli
49	7	0.9	235	4	US-09-580-235-6	Sequence 4, Appli
50	7	0.9	235	4	US-09-580-235-8	Sequence 8, Appli
51	7	0.9	235	4	US-09-580-181-2	Sequence 2, Appli
52	7	0.9	235	4	US-09-580-181-4	Sequence 4, Appli
53	7	0.9	235	4	US-09-580-181-6	Sequence 6, Appli
54	7	0.9	235	4	US-09-580-181-8	Sequence 8, Appli
55	7	0.9	235	4	US-09-102-530-2	Sequence 2, Appli
56	7	0.9	235	4	US-09-102-530-4	Sequence 4, Appli
57	7	0.9	235	4	US-09-102-530-6	Sequence 6, Appli
58	7	0.9	235	4	US-09-102-530-8	Sequence 8, Appli
59	7	0.9	235	4	US-09-620-405B-487	Sequence 487, App
60	7	0.9	235	4	US-09-604-287A-487	Sequence 487, App
61	7	0.9	235	4	US-09-834-759-487	Sequence 487, App
62	7	0.9	240	4	US-09-107-532A-6634	Sequence 6634, Ap
63	7	0.9	245	4	US-09-252-991A-29404	Sequence 29404, A
64	7	0.9	255	4	US-09-489-039A-13663	Sequence 13663, A
65	7	0.9	257	4	US-09-579-845-10	Sequence 10, Appl
66	7	0.9	261	1	US-08-622-353-2	Sequence 2, Appli
67	7	0.9	261	2	US-08-622-352A-2	Sequence 2, Appli
68	7	0.9	261	3	US-08-826-390-2	Sequence 2, Appli
69	7	0.9	298	4	US-09-252-991A-21944	Sequence 21944, A
70	7	0.9	305	4	US-09-252-991A-17702	Sequence 17702, A
71	7	0.9	312	4	US-09-252-991A-31029	Sequence 31029, A
72	7	0.9	325	4	US-09-252-991A-23807	Sequence 23807, A
73	7	0.9	364	4	US-09-134-000C-4275	Sequence 4275, Ap
74	7	0.9	381	4	US-09-198-452A-745	Sequence 745, App
75	7	0.9	383	4	US-09-459-749D-17	Sequence 17, Appl
76	7	0.9	390	3	US-08-861-774E-82	Sequence 82, Appl
77	7	0.9	406	4	US-09-721-870-46	Sequence 46, Appl
78	7	0.9	406	4	US-08-311-731A-12	Sequence 12, Appl
79	7	0.9	425	4	US-09-721-870-109	Sequence 109, App
80	7	0.9	434	4	US-09-489-039A-11674	Sequence 11674, A
81	7	0.9	447	4	US-09-252-991A-22113	Sequence 22113, A
82	7	0.9	461	1	US-08-385-229-2	Sequence 2, Appli
83	7	0.9	461	2	US-08-650-000-2	Sequence 2, Appli
84	7	0.9	461	3	US-09-042-785A-7	Sequence 7, Appli
85	7	0.9	461	3	US-08-477-347-3	Sequence 3, Appli
86	7	0.9	461	3	US-09-006-353A-4	Sequence 4, Appli
87	7	0.9	461	3	US-08-476-862-2	Sequence 2, Appli
88	7	0.9	461	4	US-09-573-986-4	Sequence 4, Appli
89	7	0.9	461	4	US-08-406-824A-2	Sequence 2, Appli
90	7	0.9	461	4	US-09-800-909-2	Sequence 2, Appli

Fri Oct 1 15:35:50 2004

us-09-765-271-56.oli.rsp

Page 16

Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 GVAVPHG 309  
Db 59 GVAVPHG 65

RESULT 24  
MUTT STRAM  
ID MUTT STRAM STANDARD; PRT; 154 AA.  
AC F32091;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE MUT-like protein (ORF154).  
OS Streptomyces ambifaciens.  
OG Plasmid pSAM2.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1889;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 23877;  
RX MEDLINE=95020551; PubMed=7934842;  
RA Hagege J., Pernodet J.L., Friedmann A., Guerinneau M.;  
RT "Mode and origin of replication of pSAM2, a conjugative integrating  
element of Streptomyces ambifaciens.";  
RL Mol. Microbiol. 10:799-812(1993).  
CC -!- SIMILARITY: Belongs to the NUDIX hydrolase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; Z19590; CAA79638.1; --  
DR PIR; S39873; S39873.  
DR HSSP; P08337; ITUM.  
DR InterPro; IPR000086; NUDIX\_hydrolase.  
DR Pfam; PF00293; NUDIX; 1.  
DR PRINTS; PR00502; NUDIXFAMILY.  
DR PROSITE; PS00893; NUDIX; 1.  
KW Plasmid; Hydrolase.  
FT DOMAIN 48 69 NUDIX BOX.  
SQ SEQUENCE 154 AA; 16689 MW; 8EC2EE99D272D656 CRC64;  
Query Match 0.9%; Score 7; DB 1; Length 154;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LTPDEVS 39  
Db 114 LTPDEVS 120

RESULT 25  
GREA STAM  
ID GREA STAM STANDARD; PRT; 158 AA.  
AC Q99TN9;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Transcription elongation factor greA (Transcript cleavage factor  
DE greA).  
GN GREA OR SAV1610 OR SA1438 OR MW1560.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain MW2).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

NCBI\_TaxID=158878, 158879, 196620;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=MU50 / ATCC 700699, and N315;  
RC MEDLINE=21311952; PubMed=11418146;  
RX Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kato C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Fuyua K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus";  
RL Lancet 357:1225-1240(2001).  
RN [2]  
SEQUENCE FROM N.A.  
RP STRAIN=MW2;  
RC MEDLINE=22040717; PubMed=12044378;  
RX Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-  
RT acquired MRSA";  
RL Lancet 359:1819-1827(2002).  
CC -!- FUNCTION: Necessary for efficient RNA polymerase transcription  
CC elongation past template-encoded arresting sites. The arresting  
CC sites in DNA have the property of trapping a certain fraction of  
CC elongating RNA polymerases that pass through, resulting in locked  
CC ternary complexes. Cleavage of the nascent transcript by cleavage  
CC factors such as greA or greB allows the resumption of elongation  
CC from the new 3' terminus. GreA releases sequences of 2 to 3  
CC nucleotides (by similarity).  
CC -!- SIMILARITY: Belongs to the greA/greB family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF003362; BAB57772.1; --  
DR EMBL; AF003362; BAB57772.1; --  
DR EMBL; AF004827; BAB95425.1; --  
DR PIR; A89943; A89943.  
DR HSSP; P21346; 1GRJ.  
DR SWISS-2DPAGE; Q99TN9; STAA.  
DR HAMAP; MF 00105; 1.  
DR InterPro; IPR006359; GreA.  
DR InterPro; IPR001437; GreA\_GreB.  
DR Pfam; PF01272; GreA\_GreB; 1.  
DR Pfam; PF03449; GreA\_GreB; 1.  
DR ProDom; PD004918; GreA\_GreB; 1.  
DR TIGRFAMs; TIGR01462; greA; 1.  
DR PROSITE; PS00829; GREAB; 1.  
DR PROSITE; PS00830; GREAB\_2; 1.  
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.  
FT DOMAIN 4 76 COILED COIL (POTENTIAL).  
SQ SEQUENCE 158 AA; 17743 MW; EC3B0F0E6238A107 CRC64;  
Query Match 0.9%; Score 7; DB 1; Length 158;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 IIEDTGD 182  
Db 77 IIEDTGD 83

Search completed: October 1, 2004, 07:32:52  
Job time : 24 secs



```

FT ACT SITE 119 119 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC... ) (PARTIAL).
SQ SEQUENCE 124 AA; 13711 MW; 9435EF532420F852 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SNPSSVS 791
Db 15 SNPSSVS 21

RESULT 22
Y082 RICPR STANDARD; PRT; 143 AA.
AC Q9ZE65;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RF082.
GN RF082.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AJ235270; CAAL4552.1; -
PIR; A71717; A71717.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 16939 MW; 57637C7A62FE9F10 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 NRQKEH 139
Db 106 NRQKEH 112

RESULT 23
PTVA ECOLI STANDARD; PRT; 148 AA.
AC P32155; P76776;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PTS system, fructose-like-1 IIA component (Phosphotransferase enzyme
DE II, A component) (EC 2.7.1.69).
GN FRVA OR B3900.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=93374854; PubMed=8396120;
RA Moralejo P., Egan S.M., Hidalgo E.F., Aguilar J.;
RT "Sequencing and characterization of a gene cluster encoding the
RT enzymes for L-rhamnose metabolism in Escherichia coli.";
RL J. Bacteriol. 175:5585-5594(1993).

[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes.";
RL Nucleic Acids Res. 21:3391-3398(1993).

[3]
RN REVISIONS TO 81 AND 104-108.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).

[4]
RN DISCUSSION OF SEQUENCE.
RP MEDLINE=94290319; PubMed=8019415;
RX Reizer J., Michotey V., Reizer A., Saier M.H. Jr.;
RT "Novel phosphotransferase system genes revealed by bacterial genome
RT analysis: unique, putative fructose- and glucoside-specific
RT systems.";
RL Protein Sci. 3:440-450(1994).

CC -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
CC sugar phosphotransferase system (PTS), a major carbohydrate active
CC -transport system. The IID domains contain the sugar binding site
CC and the transmembrane channel; the IIA domain contains the primary
CC phosphorylation site (the donor is phospho-HPr); IIA transfers its
CC phosphoryl group to the IIB domain which finally transfers it to
CC the sugar.
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Contains 1 PTS EIIA domain.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; X60472; CA433004.1; -
EMBL; L19201; AAB03033.2; -
PIR; D48649; D48649.
EcGene; EG11864; frvA.
DR InterPro; IPR002178; PTS_EIIA_2.
DR InterPro; IPR004715; PTSIIA fruc.
DR Pfam; PF00359; PTS_EIIA_2; 1.
DR ProDom; PD001699; PTS_EIIA_2; 1.
DR TIGRPFAMs; TIGR00848; frvA; 1.
DR PROSITE; PS00372; PTS_EIIA_2; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Complete proteome.
FT MOD_RES 64 64 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 81 81 A->G (IN REF. 2).
FT CONFLICT 104 107 QSGE -> KXZ (IN REF. 2).
SQ SEQUENCE 148 AA; 16093 MW; 04AE87B9084BA1C6 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 148;

```

DR EMBL; D90778; BAA15013.1; ALT INIT.  
 DR EMBL; D90779; BAA15019.1; ALT INIT.  
 DR EMBL; D90838; BAA15822.1; ALT INIT.  
 DR EMBL; D90850; BAA16005.1; ALT INIT.  
 DR EMBL; D90851; BAA16013.1; ALT INIT.  
 DR EMBL; D90852; BAA16036.1; ALT INIT.  
 DR EcoGene; EG40003; insc.  
 DR InterPro; IPR002514; Transposase 8.  
 DR Pfam; PF01527; Transposase 8; 1\_  
 KW Transposable element; Transposition; DNA-binding; DNA recombination;  
 KW Complete proteome.  
 FT VARIANT 34 34 L -> F (IN B1997).  
 SQ SEQUENCE 121 AA; 13452 MW; 59431E5C452E067A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 202 ASELAAA 208  
 DB 70 ASELAAA 76

RESULT 20  
 INSC SHIFL STANDARD; PRT; 121 AA.  
 AC P59444;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Transposase insc for insertion element IS2.  
 GN (INSC1 OR SF0245) AND (INSC2 OR SF0879) AND (INSC3 OR SF0933) AND  
 GN (INSC4 OR SF0960) AND (INSC5 OR SF1054) AND (INSC6 OR SF1165) AND  
 GN (INSC7 OR SF1343) AND (INSC8 OR SF1463) AND (INSC9 OR SF1587) AND  
 GN (INSC10 OR SF2011) AND (INSC11 OR SF2615) AND  
 GN (INSC12 OR SF2694) AND (INSC13 OR SF2873) AND  
 GN (INSC14 OR SF2984) AND (INSC15 OR SF3431) AND  
 GN (INSC16 OR SF3512) AND (INSC17 OR SF3805) AND  
 GN (INSC18 OR SF3873) AND (INSC19 OR SF3988) AND  
 GN (INSC20 OR SF4097) AND (INSC21 OR SF4185).  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OC NCBI\_TaxID=623;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 through comparison with genomes of Escherichia coli K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 CC -!- FUNCTION: Involved in the transposition of the insertion  
 CC sequence IS2 (By similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE015060; AAN41906.1; ALT INIT.  
 DR EMBL; AE015116; AAN42511.1; ALT INIT.  
 DR EMBL; AE015123; AAN42562.1; ALT INIT.  
 DR EMBL; AE015125; AAN42588.1; ALT INIT.  
 DR EMBL; AE015133; AAN42676.1; ALT INIT.  
 DR EMBL; AE015143; AAN42781.1; ALT INIT.

DR EMBL; AE015160; AAN42947.1; ALT INIT.  
 DR EMBL; AE015170; AAN43060.1; ALT INIT.  
 DR EMBL; AE015180; AAN43173.1; ALT INIT.  
 DR EMBL; AE015218; AAN43556.1; ALT INIT.  
 DR EMBL; AE015277; AAN44112.1; ALT INIT.  
 DR EMBL; AE015284; AAN44187.1; ALT INIT.  
 DR EMBL; AE015301; AAN44359.1; ALT INIT.  
 DR EMBL; AE015311; AAN44465.1; ALT INIT.  
 DR EMBL; AE015351; AAN44892.1; ALT INIT.  
 DR EMBL; AE015359; AAN44970.1; ALT INIT.  
 DR EMBL; AE015387; AAN45245.1; ALT INIT.  
 DR EMBL; AE015394; AAN45310.1; ALT INIT.  
 DR EMBL; AE015405; AAN45422.1; ALT INIT.  
 DR EMBL; AE015418; AAN45522.1; ALT INIT.  
 DR EMBL; AE015427; AAN45606.1; ALT INIT.  
 DR InterPro; IPR002514; Transposase 8.  
 DR Pfam; PF01527; Transposase 8; 1\_  
 KW Transposable element; Transposition; DNA-binding; DNA recombination;  
 KW Complete proteome.  
 SQ SEQUENCE 121 AA; 13452 MW; 59431E5C452E067A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 ASELAAA 208  
 DB 70 ASELAAA 76

RESULT 21  
 RNP ANTAM STANDARD; PRT; 124 AA.  
 AC P00668;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A).  
 GN RNASE1 OR ENS1.  
 OS Antilocapra americana (Pronghorn).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Antilocapridae; Antilocapra.  
 OC NCBI\_TaxID=9891;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=80075014; PubMed=513141;  
 RA Beintema J.J., Gastra W., Munnikema J.;  
 RT "Primary structure of pronghorn pancreatic ribonuclease: close  
 RT relationship between giraffe and pronghorn.";  
 RL J. Mol. Evol. 13:305-316(1979).  
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-  
 CC phosphates and 3'-phosphooligonucleotides ending in C-P or U-P  
 CC with 2',3'-cyclic phosphate intermediates.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Pancreas.  
 CC -!- SIMILARITY: Belongs to the pancreatic ribonuclease family.  
 DR PIR; A00813; NRPNH.  
 DR HSP; P00656; LRBG.  
 DR InterPro; IPR001427; RNaseA.  
 DR Pfam; PF00074; RNaseA; 1.  
 DR PRINTS; PR00794; RIBONUCLEASE.  
 DR ProDom; PD000535; RNaseA; 1.  
 DR SMART; SM00092; RNase\_Pc; 1.  
 DR PROSITE; PS00127; RNASE\_PANCREATIC; 1.  
 DR Hydrolase; Nuclease; Endonuclease; Glycoprotein.  
 FT DISULFID 26 84 BY SIMILARITY.  
 FT DISULFID 40 95 BY SIMILARITY.  
 FT DISULFID 58 110 BY SIMILARITY.  
 FT DISULFID 65 72 BY SIMILARITY.  
 FT ACT\_SITE 12 12 BY SIMILARITY.  
 FT ACT\_SITE 41 41 BY SIMILARITY.

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ribonuclease P protein component (EC 3.1.26.5) (RNaseP protein) (RNase  
DE P protein) (Protein C5).  
GN RNPA OR SE2418.  
OS Staphylococcus epidermidis.  
CC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RX PubMed=12950922;  
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,  
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;  
RT "Genome-based analysis of virulence genes in a non-biofilm-forming  
RT Staphylococcus epidermidis strain (ATCC 12228).";  
RL Mol. Microbiol. 49:1577-1593(2003).  
CC  
CC -!- FUNCTION: RNaseP catalyzes the removal of the 5'-leader sequence  
CC from pre-tRNA to produce the mature 5'terminus. It can also cleave  
CC other RNA substrates such as 4.5S RNA. The protein component plays  
CC an auxiliary but essential role in vivo by binding to the 5'-  
CC leader sequence and broadening the substrate specificity of the  
CC ribozyme (By similarity).  
CC  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-  
CC extra-nucleotide from tRNA precursor.  
CC  
CC -!- SUBUNIT: Consists of a catalytic RNA component (M1 or rnpB) and a  
CC protein subunit (By similarity).  
CC  
CC -!- SIMILARITY: Belongs to the rnpA family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE016752; AAC06061.1; -.  
DR HAMAP; MF 00227; -; 1.  
DR InterPro; IPR00100; Ribonuclease P.  
DR Pfam; PF00825; Ribonuclease P; 1.  
DR ProDom; PD003629; Ribonuclease P; 1.  
DR TIGRFAMs; TIGR00188; rnpA; 1.  
DR PROSITE; PS00648; RIBONUCLEASE\_P; 1.  
KW Hydrolase; Nuclease; Endonuclease; tRNA processing; RNA-binding;  
KW Complete proteome.  
SQ SEQUENCE 115 AA; 13484 MW; 106B2592C8400F18 CRC64;  
  
Query Match 0.9%; Score 7; DB 1; Length 115;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 460 RNSDFQA 466  
| | | | |  
Db 9 RNSDFQA 15  
  
RESULT 19  
ID INSC ECOLI STANDARD; PRT; 121 AA.  
AC P19776; O07989; O08018; O08019; P76357; P77346;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transposase insC for insertion element IS2A/D/F/H/I/K.  
GN (INSC1 OR B0360) AND (INSC2 OR B1403) AND (INSC3 OR B1997) AND  
GN (INSC4 OR B2861) AND (INSC5 OR B3044) AND (INSC6 OR B4272).  
OS Escherichia coli  
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
CC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=K12;  
RX MEDLINE=88137965; PubMed=2830172;  
RA Ronecker H.J., Rak B.;  
RT "Genetic organization of insertion element IS2 based on a revised  
RT nucleotide sequence.";  
RL Gene 59:291-296(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=95334362; PubMed=7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RA Blattner F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes";  
RL Nucleic Acids Res. 23:2105-2119(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,  
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,  
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
RA Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,  
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,  
RA Yamamoto Y., Horiuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 40.1-50.0 min region on the linkage map.";  
RL DNA Res. 3:379-392(1996).  
CC  
CC -!- FUNCTION: Involved in the transposition of the insertion sequence  
CC IS2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; V00279; CAA23542.1; -.  
DR EMBL; U14003; AAA97168.1; ALT\_INIT.  
DR EMBL; U28377; AAA69212.1; ALT\_INIT.  
DR EMBL; U28375; AAA83043.1; ALT\_INIT.  
DR EMBL; AE000143; AAC73463.1; ALT\_INIT.  
DR EMBL; AE000237; AAC73463.1; ALT\_INIT.  
DR EMBL; AE000231; AAC74485.1; ALT\_INIT.  
DR EMBL; AE000291; AAC75058.1; ALT\_INIT.  
DR EMBL; AE000369; AAC75900.1; -.  
DR EMBL; AE000386; AAC76080.1; -.  
DR EMBL; AE000498; AAC77228.1; ALT\_INIT.

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; Z98980; CAB11722.1; --  
DR EMBL; AB015167; BAA28750.1; --  
DR EMBL; AF192764; AAF19628.1; --  
DR EMBL; AF121275; AAD25391.1; --  
DR PIR; T38824; T38824.  
DR HSP; P35754; 1JHB.  
DR GeneDB\_SPombe; SPAC4F10.20; --  
DR InterPro; IPR002109; Glutaredoxin.  
DR InterPro; IPR006663; Thioredox\_dom2.  
DR Pfam; PF00462; glutaredoxin; 1.  
DR PRINTS; PR00160; GLUTAREDOXIN.  
DR PROSITE; PS00195; GLUTAREDOXIN; 1.  
KW Redox-active center; Electron transport.  
FT DISULFID 25 28 REDOX-ACTIVE (BY SIMILARITY).  
FT CONFLICT 52 52 N -> D (IN REF. 3).  
FT CONFLICT 101 AA; 11261 MW; 30557E19BF33E9BB CRC64;  
SQ SEQUENCE 101 AA; 11261 MW; 30557E19BF33E9BB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 101;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 NSDFQAL 467  
DB 82 NSDFQAL 88

RESULT 16  
RS16\_UREPA  
ID RS16\_UREPA STANDARD; PRT; 101 AA.  
AC Q9PP51;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 30S ribosomal protein S16.  
GN RSP OR RPS16 OR UUS68.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Serovar 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Casell G.H.;  
RA "The complete sequence of the mucosal pathogen Ureaplasma  
urealyticum";  
RT Nature 407:757-762 (2000).  
RL Nature 407:757-762 (2000).  
CC -!- SIMILARITY: Belongs to the S16P family of ribosomal proteins.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; AE002155; AAF30982.1; --  
DR HSP; P80379; 1EMW.  
DR HAMAP; MF 00385; --; 1.  
DR InterPro; IPR000307; Ribosomal\_S16.  
DR Pfam; PF00886; Ribosomal\_S16; 1.  
DR ProDom; PD003791; Ribosomal\_S16; 1.  
DR TIGRFAMS; TIGR00002; S16; 1.  
DR PROSITE; PS00732; RIBOSOMAL\_S16; FALSE\_NEG.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 101 AA; 11297 MW; FE91E1AB36CBF45 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 101;

Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 PSETVKX 410  
DB 66 PSETVKX 72

RESULT 17  
YQCC\_HAEIN STANDARD; PRT; 106 AA.  
ID YQCC\_HAEIN  
AC Q57152; O05061;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein H11436.  
GN H11436.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd";  
RL Science 269:496-512 (1995).  
CC -!- SIMILARITY: TO THE N-TERMINAL OF E.CAROTOVORA EXOENZYME REGULATION  
CC REGULON ORF1. THE C-TERMINAL PART IS COLINEAR WITH YQCB.  
CC -!- SIMILARITY: STRONG, TO E.COLI YQCC.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; U32822; AAC23085.1; --  
DR PIR; I64171; I64171.  
DR TIGR; H11436; --  
DR InterPro; IPR007384; DUF446.  
DR Pfam; PF04287; DUF446; 1.  
DR PIRSF; PIRSF06257; UCP006257; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 106 AA; 12273 MW; 0955920EBD63228C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 106;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 AAEAFLS 213  
DB 28 AAEAFLS 34

RESULT 18  
RNPA\_STAEP STANDARD; PRT; 115 AA.  
ID RNPA\_STAEP  
AC Q8CMN4;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)

Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL J. Bacteriol. 184:5479-5490(2002).  
 [7]  
 RP SEQUENCE OF 1-15.  
 RX MEDLINE=92176646; PubMed=1371791;  
 RA Barnes P.F., Mehra V., Rivoire B., Fong S.J., Brennan P.J.,  
 RA Voegtline M.S., Minden P., Houghten R.A., Bloom B.R., Modlin R.L.;  
 RA "Immunoreactivity of a 10-kDa antigen of Mycobacterium tuberculosis.";  
 J. Immunol. 148:1835-1840(1992).  
 CC -1- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses  
 CC the ATPase activity of the latter.  
 CC -1- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the groES chaperonin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X60350; CAA42908.1; -;  
 DR EMBL; M25258; AAA25340.1; -;  
 DR EMBL; X13739; CAA32003.1; -;  
 DR EMBL; Z77165; CAB01005.1; -;  
 DR EMBL; AE007158; AAK47865.1; -;  
 DR PIR; S01381; BVMYBA  
 DR PDB; 1HX5; 08-AUG-01.  
 DR PDB; 1P3H; 15-JUL-03.  
 DR TIGR; MT3527; -;  
 DR TuberculList; RV3418C; -;  
 DR HAMAP; MF\_00580; -; 1.  
 DR InterPro; IPR001476; Chaprnnin\_Cpn10.  
 DR Pfam; PF00166; cpn10; 1.  
 DR PRINTS; PR00297; CHAPERINN10.  
 DR PRODOM; PD000566; Chaprnnin\_Cpn10; 1.  
 DR PROSITE; PS00681; CHAPERONINS CPN10; 1.  
 KW Chaperone; Antigen; Heat shock; Complete proteome; 3D-structure.  
 FT INIT MET 0  
 SQ SEQUENCE 99 AA; 10673 MW; 1DD128E75CF19AF7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 GEEKRIPL 611  
 |||||  
 DB 54 GEEKRIPL 60

RESULT 15  
 GLRL\_SCHPO STANDARD; PRT; 101 AA.  
 AC GUR1 SCHPO Q9US58;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutaredoxin 1  
 GN GRX1 OR SPAC4F10.20.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawamukai M.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Kim H.-G., Cho Y.-W., Park E.-H., Lim C.-J.;  
 RT "Characterization of cDNA encoding thioltransferase (glutaredoxin)  
 RT from Schizosaccharomyces pombe.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Cho Y.-W., Kim H.-G., Lim C.-J.;  
 RT "Isolation and expression of the genomic DNA encoding thioltransferase  
 RT (Glutaredoxin) from Schizosaccharomyces pombe.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Thode G.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: THE DISULFIDE BOND FUNCTIONS AS AN ELECTRON CARRIER IN  
 CC THE GLUTATHIONE-DEPENDENT SYNTHESIS OF DEOXYRIBONUCLEOTIDES BY THE  
 CC ENZYME RIBONUCLEOTIDE REDUCTASE. IN ADDITION, IT IS ALSO INVOLVED  
 CC IN REDUCING SOME DISULFIDES IN A COUPLED SYSTEM WITH GLUTATHIONE  
 CC REDUCTASE. THIOLTRANSFERASE CATALYZES CELLULAR THIOL-DISULFIDE  
 CC TRANSYDROGENATION REACTIONS. IT TRANSFERS REDUCING EQUIVALENTS  
 CC TO CYTOSOLIC PROTEIN AND NONPROTEIN DISULFIDES (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the glutaredoxin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ248285; CAB49786.1; -;  
 DR PIR; A75134; A75134.  
 DR InterPro; IPR008203; DUF104.  
 DR Pfam; PF01954; DUF104; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 70 AA; 8009 MW; 4BDE6021B/DFP8032 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 478 STNKEKL 484 ✓  
 DB 34 STNKEKL 40

## RESULT 13

CH10\_MYCBO STANDARD; PRT; 99 AA.  
 ID CH10\_MYCBO  
 AC P15020;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 10 kDa chaperonin (Protein Cpn10) (groES protein) (Immunogenic protein  
 DE MPB57).  
 GN GROS OR GROES OR MOPB OR MB3452C.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.  
 RC STRAIN=BCG;  
 RX MEDLINE=89052868; PubMed=3056744;  
 RA Yamaguchi R., Matsuo K., Yamazaki A., Nagai S., Terasaka K.,  
 RA Yamada T.;  
 RT "Immunogenic protein MPB57 from Mycobacterium bovis BCG: molecular  
 RT cloning, nucleotide sequence and expression.";  
 RL FEBS Lett. 240:115-117(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 CC -!- FUNCTION: Binds to Cpn60 in the presence of Mg-ATP and suppresses  
 CC the ATPase activity of the latter.  
 CC -!- SUBUNIT: Heptamer of 7 subunits arranged in a ring (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the groES chaperonin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X13970; CAA32149.1; -;  
 DR EMBL; M35389; AAA25365.1; -;  
 DR EMBL; BX248346; CAD95639.1; -;  
 DR PIR; S01784; BVMY7B.

DR HSSP; P05380; 1AON.  
 DR HAMAP; MF 00580; -; 1.  
 DR InterPro; IPR001476; Chaprinin\_Cpn10.  
 DR Pfam; PF00166; cpn10; 1.  
 DR PRINTS; PR00297; CHAPERONIN10.  
 DR ProDom; PD00566; Chaprinin\_Cpn10; 1.  
 DR PROSITE; PS00681; CHAPERONINS\_CPN10; 1.  
 KW Chaperone; Antigen; Heat shock; Complete proteome.  
 FT INIT MET 0 0  
 FT CONFLICT 94 98 LAVVS -> VGRRF (IN REF. 1).  
 SQ SEQUENCE 99 AA; 10673 MW; 1DD128E75CF19AF7 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 605 GEKRIPL 611  
 DB 54 GEKRIPL 60

## RESULT 14

CH10\_MYCTU STANDARD; PRT; 99 AA.  
 ID CH10\_MYCTU  
 AC P09621;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 10 kDa chaperonin (Protein Cpn10) (groES protein) (BCG-A heat shock  
 DE protein) (10 kDa antigen).  
 GN GROS OR GROES OR MOPB OR CPN10 OR RV3418C OR MT3527 OR MTCV78.11.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=89016584; PubMed=2902558;  
 RA Baird P.N., Hall L.M., Coates A.R.M.;  
 RT "A major antigen from Mycobacterium tuberculosis which is homologous  
 RT to the heat shock proteins groES from E. coli and the htpA gene  
 RT product of *Coxiella burnetii*.";  
 RL Nucleic Acids Res. 16:9047-9047(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=90095443; PubMed=2480990;  
 RA Baird P.N., Hall L.M.C., Coates A.R.M.;  
 RT "Cloning and sequence analysis of the 10 kDa antigen gene of  
 RT Mycobacterium tuberculosis.";  
 RL J. Gen. Microbiol. 135:931-939(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Erdmann;  
 RX MEDLINE=89160258; PubMed=2564178;  
 RA Shinnick T.M., Plikaytis B.P., Hyshe A.D., van Landingham R.M.,  
 RA Walker L.L.;  
 RT "The Mycobacterium tuberculosis BCG-a protein has homology with the  
 RT Escherichia coli GroES protein.";  
 RL Nucleic Acids Res. 17:1254-1254(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Erdmann;  
 RX MEDLINE=93219332; PubMed=7681982;  
 RA Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;  
 RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

FT LIPID 871 871 Omega-hydroxyceramide glutamate ester  
 FT LIPID 1607 1607 (Probable).  
 FT LIPID 1607 1607 Omega-hydroxyceramide glutamate ester  
 FT LIPID 1607 1607 (Probable).  
 SQ SEQUENCE 2033 AA; 231616 MW; B8DC6E2B52221938 CRC64;  
 Query Match 1.0%; Score 8; DB 1; Length 2033;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 731 LKEAEVLL 738  
 DB 87 LKEAEVLL 94  
 RESULT 11  
 EVPL MOUSE STANDARD; PRT; 2035 AA.  
 AC Q9D952;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Envoplakin (p210) (210 kDa cornified envelope precursor).  
 GN EVPL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=20347896; PubMed=10747979;  
 RA Maatta A., Ruhnberg C., Watt F.M.;  
 RT "Structure and regulation of the envoplakin gene.";  
 RL J. Biol. Chem. 275:19857-19865(2000).  
 RN [2]  
 RP SEQUENCE OF 1860-2035 FROM N.A.  
 RC STRAIN=C57BL/6J; Tissue=Pancreas;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pescio G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mombauts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -!- FUNCTION: Component of the cornified envelope of keratinocytes.  
 CC May link the cornified envelope to desmosomes and intermediate  
 CC filaments.  
 CC -!- SUBUNIT: May form a homodimer or a heterodimer with PPL.  
 CC -!- SUBCELLULAR LOCATION: COLocalized WITH DSP AT DESMOSOMES AND ALONG  
 CC INTERMEDIATE FILAMENTS (BY SIMILARITY).  
 CC -!- SIMILARITY: Contains 7 plectrin repeats.  
 CC -!- SIMILARITY: Contains 1 spectrin repeat.  
 CC -!- SIMILARITY: Belongs to the plectrin repeat.  
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----

DR EMBL; AJ309317; CAC38864.2; JOINED.  
 DR EMBL; AJ319607; CAC38864.2; JOINED.  
 DR EMBL; AJ319608; CAC38864.2; JOINED.  
 DR EMBL; AJ319609; CAC38864.2; JOINED.  
 DR EMBL; AJ319610; CAC38864.2; JOINED.  
 DR EMBL; AJ319611; CAC38864.2; JOINED.  
 DR EMBL; AJ319612; CAC38864.2; JOINED.  
 DR EMBL; AJ319613; CAC38864.2; JOINED.  
 DR EMBL; AK007353; NOT\_ANNOTATED\_CDS.  
 DR MGD; MGI:107507; Evpl; EMBL repeat.  
 DR InterPro; IPR001101; Plectrin\_repeat.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00681; Plectrin; 3.  
 DR Pfam; PF00435; spectrin; 2.  
 DR SMART; SM00250; PLEC; 8.  
 KW Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.  
 FT DOMAIN 1 841 GLOBULAR 1.  
 FT DOMAIN 842 1674 CENTRAL FIBROUS ROD DOMAIN.  
 FT DOMAIN 1675 2035 GLOBULAR 2.  
 FT DOMAIN 12 28 4 X 4 AA TANDEM REPEATS OF K-G-S-P.  
 FT REPEAT 229 330 SPECTRIN.  
 FT REPEAT 842 1664 COILED COIL (POTENTIAL).  
 FT REPEAT 1186 1227 PLECTIN 1.  
 FT REPEAT 1679 1714 PLECTIN 2.  
 FT REPEAT 1819 1856 PLECTIN 3.  
 FT REPEAT 1857 1894 PLECTIN 4.  
 FT REPEAT 1895 1932 PLECTIN 5.  
 FT REPEAT 1933 1970 PLECTIN 6.  
 FT REPEAT 1971 2008 PLECTIN 7.  
 FT REPEAT 1861 1864 RPRQ -> EAQA (IN REF. 2).  
 FT CONFLICT 2035 AA; 232317 MW; EB41D9CAED9641F CRC64;  
 SQ SEQUENCE

Query Match 1.0%; Score 8; DB 1; Length 2035;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 LKEAEVLL 738  
 DB 87 LKEAEVLL 94

RESULT 12  
 Y872 PYRAB STANDARD; PRT; 70 AA.  
 AC Q9V0B8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Hypothetical UPF0165 protein PYRAB08720.  
 GN PYRAB08720 ORF0165 protein PYRAB08720.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OC NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GES / Orsay;  
 RX MEDLINE=22511545; PubMed=12622808;  
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
 RA Poch O., Priet D., Querellou J., Ripp R., Thierry J.-C.,  
 RA van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
 RT "An integrated analysis of the genome of the hyperthermophilic  
 RT archaeon Pyrococcus abyssi.";  
 RL Mol. Microbiol. 47:1495-1512(2003).  
 CC -!- SIMILARITY: Belongs to the UPF0165 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Deicher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL J. Bacteriol. 184:5479-5490(2002).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND  
 CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).  
 CC -1- SIMILARITY: Belongs to the SMC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ14609; CAC93884.1; -- INIT.  
 CC EMBL; Z74697; CAA98982.1; ALT\_INIT.  
 CC TIGR; MT2990; --  
 CC TubercuList; Rv2922c; --  
 CC InterPro; IPR003439; ABC transporter.  
 CC InterPro; IPR003405; SMC\_C.  
 CC InterPro; IPR003395; SMC\_N.  
 CC Pfam; PF02483; SMC\_C; 1.  
 CC Pfam; PF02463; SMC\_N; 1.  
 CC ATP-binding; Coiled coil; Complete proteome.  
 KW NP BIND 31 38 ATP (POTENTIAL).  
 FT DOMAIN 167 289 COILED COIL (POTENTIAL).  
 FT DOMAIN 330 499 COILED COIL (POTENTIAL).  
 FT DOMAIN 659 842 COILED COIL (POTENTIAL).  
 FT DOMAIN 899 929 COILED COIL (POTENTIAL).  
 FT DOMAIN 979 1038 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1205 AA; 130637 MW; A3B2A813B58EACF3 CRC64;  
 Query Match 1.0%; Score 8; DB 1; Length 1205;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 203 SELAAAEA 210  
 Db 672 SELAAAEA 679  
 RESULT 10  
 ID EVPL HUMAN STANDARD; PRT; 2033 AA.  
 AC Q928T7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Envoplakin (210 kDa paraneoplastic pemphigus antigen) (p210) (210 kDa  
 DE cornified envelope precursor).  
 GN EVPL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Keratinocytes;  
 RX MEDLINE=96326676; PubMed=8707850;  
 RA Ruhrberg C., Hajibagheri M.A.N., Simon M., Dooley T.P., Watt F.M.;  
 RT "Envoplakin, a novel precursor of the cornified envelope that has  
 RT homology to desmoplakin";  
 RL J. Cell Biol. 134:715-729(1996).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99339988; PubMed=10409435;  
 RA Risk J.M., Ruhrberg C., Hennies H.-C., Mills H.S., Di Colandrea T.,  
 RA Evans K.B., Ellis A., Watt F.M., Bishop D.T., Spurr N.K.,  
 RA Stevens H.P., Leigh I.M., Reis A., Kelsell D.P., Field J.K.;  
 RT "Envoplakin, a possible candidate gene for focal NEPK/esophageal  
 RT cancer (TOC): the integration of genetic and physical maps of the TOC  
 RT region on 17q25";  
 RL Genomics 59:234-242(1999).  
 RN [3]  
 CC LIPIDATION.  
 RP MEDLINE=98316349; PubMed=9651377;  
 RA Marekov L.N., Steinert P.M.;  
 RT "Ceramides are bound to structural proteins of the human foreskin  
 RT epidermal cornified cell envelope";  
 RL J. Biol. Chem. 273:17763-17770(1998).  
 CC -1- FUNCTION: Component of the cornified envelope of keratinocytes.  
 CC May link the cornified envelope to desmosomes and intermediate  
 CC filaments.  
 CC -1- SUBUNIT: May form a homodimer or a heterodimer with PPL.  
 CC -1- SUBCELLULAR LOCATION: COLOCALIZED WITH DSP AT DESMOSES AND ALONG  
 CC INTERMEDIATE FILAMENTS.  
 CC -1- TISSUE SPECIFICITY: Exclusively expressed in stratified squamous  
 CC epithelia.  
 CC -1- INDUCTION: During differentiation of epidermal keratinocytes.  
 CC -1- PTM: Substrate of transglutaminase. Some glutamines and lysines  
 CC are cross-linked to other envoplakin molecules, and to proteins  
 CC such as keratin, desmoplakin, periplakin and involucrin, and to  
 CC lipids like omega-hydroxyceramide.  
 CC -1- SIMILARITY: Contains 7 plectrin repeats.  
 CC -1- SIMILARITY: Contains 1 spectrin repeat.  
 CC -1- SIMILARITY: Belongs to the plakin or cytolinker family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U53786; AAC64662.1; --  
 CC EMBL; U72849; AAD00186.1; --  
 CC EMBL; U72843; AAD00186.1; JOINED.  
 CC EMBL; U72845; AAD00186.1; JOINED.  
 CC EMBL; U72846; AAD00186.1; JOINED.  
 CC EMBL; U72847; AAD00186.1; JOINED.  
 CC EMBL; U72848; AAD00186.1; JOINED.  
 CC Genew; HGNC:3503; EVPL.  
 CC MIM; 601590; --  
 CC GO; GO:0009506; C:plasmodesma; TAS.  
 CC GO; GO:0005198; F:structural molecule activity; TAS.  
 CC GO; GO:0008544; P:epidermal differentiation; TAS.  
 CC InterPro; IPR001101; Plectrin\_repeat.  
 CC Pfam; PF00681; Plectrin; 3.  
 CC SMART; SM00250; PLEC; 8.  
 KW Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein;  
 KW Lipoprotein.  
 FT DOMAIN 1 841 GLOBULAR 1.  
 FT DOMAIN 842 1673 CENTRAL FIBROUS ROD DOMAIN.  
 FT DOMAIN 1674 2033 GLOBULAR 2.  
 FT DOMAIN 22 28 4 X 4 AA TANDEM REPEATS OF K-G-S-P.  
 FT REPEAT 129 330 SPECTRIN.  
 FT DOMAIN 845 1135 COILED COIL (POTENTIAL).  
 FT REPEAT 1185 1226 PLECTIN 1.  
 FT REPEAT 1678 1713 PLECTIN 2.  
 FT REPEAT 1818 1855 PLECTIN 3.  
 FT REPEAT 1856 1893 PLECTIN 4.  
 FT REPEAT 1894 1931 PLECTIN 5.  
 FT REPEAT 1932 1969 PLECTIN 6.  
 FT REPEAT 1970 2007 PLECTIN 7.  
 FT LIPID 657 657 Omega-hydroxyceramide glutamate ester  
 FT (Potential).



RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 RN [3]  
 RN PARTIAL SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=EMBRYO;  
 RX MEDLINE=98198453; PubMed=9531534;  
 RA Steneberg P., England C., Kronham J., Weaver T.A., Samakovlis C.;  
 RT "Translational readthrough in the hdc mRNA generates a novel branching  
 RT inhibitor in the *Drosophila* trachea.";  
 RL Genes Dev. 12:956-967 (1998).  
 CC -!- FUNCTION: Required for imaginal cell differentiation, may be  
 CC involved in hormonal responsiveness during metamorphosis. Involved  
 CC in an inhibitory signaling mechanism to determine the number of  
 CC cells that will form unicellular sprouts in the trachea. Regulated  
 CC by transcription factor egg. The longer hdc protein is completely  
 CC functional and the shorter protein carries some function.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: Expressed in all imaginal cells of the embryo  
 CC and larvae. Expressed in a subset of tracheal fusion cells from  
 CC stage 14 to the end of embryogenesis in metameres 2-9, lateral  
 CC trunk and ventral anastomoses.  
 CC -!- MISCELLANEOUS: Readthrough of the terminator UAA occurs between  
 CC codons for Ala-650 and His-652. Readthrough is not always  
 CC suppressed as the shorter protein is more abundant.  
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; Z50097; CAA90425.1; -;  
 CC DR EMBL; Z50097; CAB58233.1; -;  
 CC DR EMBL; AE003773; AAF57033.1; ALT\_SEQ.  
 CC DR FlyBase; FBgn0010113; hdc.  
 CC DR GO; GO:0005737; C:cytoplasm; IDA.  
 CC DR GO; GO:0007430; P:terminal branching of trachea, cytoplasmic . . .; NAS.  
 CC DR Developmental protein.  
 CC KW CHAIN 1 1080 HEADCASE PROTEIN.  
 FT

PT CHAIN 1 650 HEADCASE SHORT PROTEIN.  
 FT DOMAIN 57 66 POLY-GLY.  
 FT DOMAIN 211 218 POLY-ASN.  
 FT DOMAIN 219 227 POLY-GLY.  
 FT DOMAIN 343 350 POLY-GLN.  
 FT DOMAIN 381 395 POLY-GLN.  
 FT DOMAIN 723 769 GLN-RICH.  
 FT DOMAIN 801 815 POLY-GLN.  
 FT DOMAIN 845 854 POLY-SER.  
 FT DOMAIN 887 891 POLY-SER.  
 FT DOMAIN 965 970 POLY-SER.  
 FT DOMAIN 1030 1036 POLY-SER.  
 FT CONFLICT 85 85 H -> P (IN REF. 1).  
 FT CONFLICT 190 191 PT -> SN (IN REF. 1).  
 FT CONFLICT 226 226 A -> G (IN REF. 1).  
 FT CONFLICT 243 244 SY -> HD (IN REF. 1).  
 FT CONFLICT 279 310 SGVLTQSALATSSIRTMSPGPARQGW (IN REF. 1).  
 FT CONFLICT 342 342 P -> A (IN REF. 1).  
 FT CONFLICT 353 353 L -> V (IN REF. 1).  
 FT CONFLICT 383 383 Q -> P (IN REF. 1).  
 FT CONFLICT 432 432 D -> E (IN REF. 1).  
 FT CONFLICT 641 641 T -> S (IN REF. 1).  
 FT CONFLICT 695 695 P -> Q (IN REF. 1).  
 FT CONFLICT 852 852 S -> SS (IN REF. 1).  
 FT CONFLICT 1067 1067 A -> R (IN REF. 1).  
 SQ SEQUENCE 1080 AA; 117446 MW; 87EB144BA0D1B787 CRC64;  
 Query Match 1.0%; Score 8; DB 1; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 247 TINTNTSNN 254 ✓  
 DB 836 TINTNTSNN 843  
 |||||  
 RESULT 9  
 SMC\_MYCTU  
 ID - SMC\_MYCTU STANDARD; PRT; 1205 AA.  
 AC Q10970;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Chromosome partition protein SMC.  
 GN SMC OR RV2922C OR MT2950 OR MIC338.11C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RA Cobbe N., Heck M.M.S.;  
 RT "Phylogenetic analysis of SMC proteins.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
CC EMBL; AB000050; AAB96152.1; -
CC PIR; S73830; S73830.
CC MEROPS; S16.004; -.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_centri.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_Lon.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008288; Peptid_S16_AS.
CC InterPro; IPR001984; Peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00830; ENDOLAPASE.
CC SMART; SM00382; AAA; 1.
CC TIGRFAMS; TIGR00763; Lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
CC Hydrolase; Serine protease; Complete proteome.
CC DOMAIN 304 308
CC NP_BIND 379 386
CC ACT_SITE 702 702
CC SEQUENCE 795 AA; 90203 MW; 7E7855082060C891 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 795;
Best Local Similarity 100.0%; Pred.No.17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 ALDKLLER 473 ✓
DB 127 ALDKLLER 134

RESULT 7
Y240 ARATH
ID Y240 ARATH STANDARD; PRT; 861 AA.
AC Q22224; Q94CF2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein At2g41620.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
RX SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA

Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
Yuan S., Shinozaki K., Davis R.M., Theologis A., Ecker J.R.;
"Empirical analysis of transcriptional activity in the Arabidopsis
genome.";
RL Science 302:842-846 (2003).
RN [3]
RP IDENTIFICATION ON 2D-GELS.
RC STRAIN=cv. Columbia;
RA Sarazin B., Tonella L., Marques K., Paesano S., Chane-Pavre L.,
RA Heller M., Sanchez J.-C., Hochstrasser D.F., Thiellement H.;
RA Unpublished observations (SEP-2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR GO: 0005634; C:nucleus; TAS.
DR GO: 0003704; F:specific RNA polymerase II transcription fa. . . ; TAS.
DR GO: 0000122; P:negative regulation of transcription from P. . . ; TAS.
DR GO: 0009615; P:response to viruses; TAS.
DR InterPro: IPR001346; IRF.
DR Pfam: PF00605; IRF; 1.
DR PRINTS: PR00267; INTFRNREGFCT.
DR ProDom: PD002355; IRF; 1.
DR SMART: SM00348; IRF; 1.
DR PROSITE: PS00601; IRF; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Alternative splicing.
FT DNA BIND 13 122 TRYPTOPHAN PENTAD REPEAT
FT VARSPPLIC 1 6 MALALA -> MPVPERPAAGDSPPGTR (in isoform
FT D).
FT VARSPPLIC 152 164 /FTID=VSP_002757.
FT VARSPPLIC 165 503 /GGPPPLAHTHA -> AQGSLGSGCTGGQ (in
FT VARSPPLIC 228 256 /FTID=VSP_002758.
FT VARSPPLIC 179 179 Missing (in isoform C).
FT CONFLICT 179 179 /FTID=VSP_002759.
FT CONFLICT 412 412 Missing (in isoform B).
FT CONFLICT 412 412 E -> K (IN REF. 2).
FT CONFLICT 412 412 Q -> R (IN REF. 3).
SQ SEQUENCE 503 AA; 54278 MW; AA6A39E0E272727C CRC64;

Query Match 1.0%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred.No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 357 PSPGPQPA 364
Db 244 PSPGPQPA 251
|||||

RESULT 6
LON_MYCPN STANDARD; PRT; 795 AA.
AC P78025;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease Ia (EC 3.4.21.53).
GN LON OR MFN332 OR MF504.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Hamelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: DEGRADES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC IN PRESENCE OF ATP. DEGRADES THE REGULATORY PROTEINS RCSA AND
CC SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC PROTEIN SUBSTRATE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC GO: 0005737; C:cytoplasm; TAS.

```

NR SEQUENCE FROM N.A.  
RC STRAIN=C6;  
RX MEDLINE=94103173; PubMed=8030224;  
RT Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;  
RT "The complete DNA sequence of Autographa californica nuclear  
RT polyhedrosis virus";  
RL Virology 202:586-605(1994).  
CC -!- SIMILARITY: TO CORRESPONDING ORF IN OPMNPV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L22858; AAA66659.1; -.  
DR PIR; E72853; E72853.  
KW Hypothetical protein.  
SQ SEQUENCE 71 AA; 8569 MW; ADFP85AC68E16DD3 CRC64;  
  
Query Match 1.0%; Score 8; DB 1; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 130 EBNRQKQ 137 ✓  
DB 20 EBNRQKQ 27  
  
RESULT 3  
SC17\_NEUCR  
ID SC17\_NEUCR STANDARD; PRT; 292 AA.  
AC Q9P6A5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable vesicular-fusion protein sc17 homolog.  
GN B1D.150.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A / FGSC 987;  
RX MEDLINE=22542210; PubMed=12655011;  
RA Manthaupt G., Montrone C., Haase D., Mewes H.-W., Aign V.,  
RA Hobeisel J.D., Fattmann B., Nyakatura G., Kempken F., Maier J.,  
RA Schulte U.;  
RT "What's in the genome of a filamentous fungus? Analysis of the  
RT Neurospora genome sequence."  
RL Nucleic Acids Res. 31:1944-1954(2003).  
CC -!- FUNCTION: Required for vesicular transport between the endoplasmic  
CC reticulum and the Golgi apparatus (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic peripheral membrane protein (By  
CC similarity).  
CC -!- SIMILARITY: Belongs to the SNAP family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL355927; CAB91264.1; -.  
DR PIR; T49361; T49361.  
DR InterPro; IPR000744; NSF\_attach.  
DR InterPro; IPR008941; TPR-like.

DR Pfam; PF02071; NSF; 2.  
DR PRINTS; PR00448; NSFATTACHMNT.  
KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack.  
SQ SEQUENCE 292 AA; 32805 MW; C9BD75DB128E19E7 CRC64;  
  
Query Match 1.0%; Score 8; DB 1; Length 292;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 150 NDGAVALA 157 ✓  
DB 150 NDGAVALA 157  
  
RESULT 4  
AAC2\_DICDI  
ID AAC2\_DICDI STANDARD; PRT; 448 AA.  
AC P14136;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE AAC-rich mRNA clone AAC11 protein (Fragment).  
DE AAC11.  
GN Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=90066348; PubMed=2511421;  
RA Shaw D.R., Richter H., Giorda R., Ohmachi T., Ennis H.L.;  
RT "Nucleotide sequences of Dictyostelium discoideum developmentally  
RT regulated cDNAs rich in (AAC) imply proteins that contain clusters of  
RT asparagine, glutamine, or threonine."  
RL Mol. Gen. Genet. 218:453-459(1989).  
CC -!- DEVELOPMENTAL STAGE: The concentration of AAC-rich mRNAs is low  
CC in dormant spores and growing cells, but increases during  
CC spore-germination and multicellular development.  
CC -!- MISCELLANEOUS: Several proteins derive from AAC-rich mRNA, which,  
CC due to a frameshift also have ACA and CAA codons and thus are  
CC Asn-, Thr- or Gln-rich.  
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16522; CAA34529.1; -.  
DR PIR; S05355; S05355.  
DR DictyBase; DDB0001892; AAC11.  
DR InterPro; IPR000637; AT\_hook.  
DR InterPro; IPR000116; Highmoblty\_IY.  
DR Pfam; PF02178; AT\_hook; 4.  
DR PRINTS; PR00929; ATHOOK.  
DR ProDom; PD005593; Highmoblty\_IY; 1.  
DR SMART; SM00384; AT\_hook; 4.  
KW DNA-binding; Repeat.  
FT NON TER 1  
FT DOMAIN 5 25 GLN-RICH.  
FT DOMAIN 115 144 ASN-RICH.  
FT DOMAIN 268 384 ASN-RICH.  
SQ SEQUENCE 448 AA; 48636 MW; D8FD80D910D99817 CRC64;  
  
Query Match 1.0%; Score 8; DB 1; Length 448;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 250 NTSNNSNT 257 ✓  
DB 378 NTSNNSNT 385

CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 519-540; 597-616  
 RP AND 857-874.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93145326; PubMed=7678780;  
 RA Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,  
 RA Tjian R.;  
 RT "Molecular cloning and functional analysis of Drosophila TAF110  
 RT reveal properties expected of coactivators.";  
 RL Cell 72:247-260(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 398-406; 520-540  
 RP AND 860-877.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93317591; PubMed=8327460;  
 RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;  
 RT "The Drosophila 110-kDa transcription factor TFIID subunit directly  
 RT interacts with the N-terminal region of the 230-kDa subunit.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng C., Drenth J., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=Berkley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarini H., Kronmiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource.";

RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -!- FUNCTION: TFIID is a multimeric protein complex that plays a  
 CC central role in mediating promoter responses to various activators  
 CC and repressors. May function as a coactivator by serving as a site  
 CC of protein-protein contact between activators like Sp1 (or btd)  
 CC and TFIID complex.  
 CC -!- SUBUNIT: TFIID is composed of TATA binding protein (TBP) and a  
 CC number of TBP-associated factors (TAFs).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P47825-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P47825-2; Sequence=VSP\_004441;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.  
 CC -!- SIMILARITY: Contains 1 TAFH/NHR1 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L06861; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; S63550; AAB27433.1; -.  
 DR EMBL; AE003528; AAF49536.1; -.  
 DR EMBL; AY069807; AAL39952.1; -.  
 DR PIR; A48184; A48184.  
 DR TRANSEAC; T02121; -.  
 DR FlyBase; FBgn0010280; Taf4.  
 DR GO; GO:0005669; C:transcription factor TFIID complex; IPI.  
 DR GO; GO:0016251; F:General RNA polymerase II transcription fac...; IPI.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IPI.  
 DR GO; GO:0006367; P:transcription initiation from Pol II promoter; IPI.  
 DR InterPro; IPR007900; TAF4.  
 DR InterPro; IPR003894; TAF\_Hom.  
 DR Pfam; PF05236; TAF4; 1.  
 DR SMART; SM00549; TAFH; 1.  
 KW Transcription regulation; Nuclear protein; Alternative splicing.  
 FT DOMAIN 293 383 TAFH/NHR1.  
 FT DOMAIN 66 82 POLY-GLN.  
 FT DOMAIN 108 111 POLY-GLN.  
 FT DOMAIN 259 265 POLY-GLN.  
 FT VARSPIC 138 207 Missing (in isoform Short).  
 FT CONFLICT 114 114 Missing (VSP\_004441).  
 FT SEQUENCE 921 AA; 99338 MW; 27E6852859872767 CRC64;  
 Query Match 1.1%; Score 9; DB 1; Length 921;  
 Best Local Similarity 100.0%; Pred.No.1.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 249 TTTNNNSNT 257 ✓  
 Db 178 TTTNNNSNT 186  
 RESULT 2  
 Y029 NPVAC STANDARD; PRT; 71 AA.  
 ID Y029 NPVAC  
 AC P41433;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 8.6 kDa protein in IAP1-SOD intergenic region.  
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=46015;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:17:54 ; Search time 19 Seconds

(without alignments)  
2181.465 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKENRVS.....KLALIKGNPSSVSKEKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.1	921	1	P47825 drosophila
2	8	1.0	71	1	P41433 atropis
3	8	1.0	292	1	Q9p6a5 neurospora
4	8	1.0	448	1	AAC2 DICDI
5	8	1.0	503	1	IR77 HUMAN
6	8	1.0	795	1	ION MYCN
7	8	1.0	861	1	Y240 ARATH
8	8	1.0	1080	1	HDC DROME
9	8	1.0	1205	1	SMC MYCTU
10	8	1.0	2033	1	EVPL HUMAN
11	8	1.0	2035	1	EVPL MOUSE
12	7	0.9	70	1	Y872 PYRAB
13	7	0.9	99	1	CH10 MYCBO
14	7	0.9	99	1	CH10 MYCTU
15	7	0.9	101	1	GLR1 SCHPO
16	7	0.9	101	1	RS16 UREPA
17	7	0.9	106	1	YQCC HAEIN
18	7	0.9	115	1	RNPA STAEP
19	7	0.9	121	1	INSC ECOLI
20	7	0.9	121	1	INSC SHIFL
21	7	0.9	124	1	RNP ANTAM
22	7	0.9	143	1	Y082 RICPR
23	7	0.9	148	1	PTVA ECOLI
24	7	0.9	154	1	MUTT STRAM
25	7	0.9	158	1	GREM STRAM
26	7	0.9	158	1	Y052 ARCFU
27	7	0.9	166	1	GC31 ARATH
28	7	0.9	174	1	HUMB DROMU
29	7	0.9	177	1	APT CHLTE
30	7	0.9	178	1	NUSG NEIMA
31	7	0.9	179	1	APT HELPU
32	7	0.9	185	1	VATE DEIRA
33	7	0.9	197	1	CYCL METEX

## RESULT 1

ID	T2D3_DROME	STANDARD;	PRT;	921 AA.
AC	P47825; P49845; Q8T9E0; Q9VUY7;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Transcription initiation factor TFIID 110 kDa subunit (TAFII-110) (110 kDa TBP-associated factor).			
GN	TAF4 OR TAF110 OR CG5444.			
OS	Drosophila melanogaster (Fruit fly).			

## ALIGNMENTS

09hou4	homo sapien
P10536	rattus norv
Q09790	schizosacch
O64628	arabidopsis
P45469	escherichia
P50038	synecococc
P50044	helicobacte
O63798	rattus norv
P97372	mus musculu
P00774	rattus norv
Q8ep66	oceanobacil
O07350	vibrio chol
P42636	biomphalari
P43689	biomphalari
P15846	trichostrom
Q39528	cladrastis
Q8nh60	homo sapien
Q8ka27	buchnera ap
Q72am8	leptospira
Q28717	oryctolagus
P45976	saccharomyc
P05987	dicylostell
P32636	agariusus bi
P20287	schistosoma
Q8ub16	agrobacteri
Q15257	homo sapien
O04385	clarkia bre
P07022	escherichia
P45550	escherichia
Q8r965	thermoanaer
Q8kae1	chlorobium
P18431	saccharomyc
Q01080	saccharomyc
Q9y972	aeropyrum p
P51026	escherichia
Q9hsc0	halobacteri
Q9y919	aeropyrum p
Q9y882	schizosacch
P25295	schizosacch
Q9X935	bacillus ha
P09516	barley yell
P23258	homo sapien
Q9z310	mus musculu
Q9nrh3	homo sapien
Q8vck3	mus musculu
P23330	xenopus lae
Q18179	caenorhabdi
Q9ck66	pasteurella
P34786	euplotes oc
P90548	euplotes oc
P53377	neurospora
P20333	homo sapien
P54403	euplotes cr
P09514	beet wester
P09515	beet wester
Q51062	neisseria g
P03173	herpes simp

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 SELAAEA 210

Db 756 SELAAEA 763

Search completed: October 1, 2004, 07:33:26  
Job time : 33 secs

C;Genetics:  
 A;Gene: lon; F10orf795  
 A;Genetic code: SGC3  
 C;Superfamily: ATP-dependent Lon protease  
 F;Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine  
 F;379-386/Region: nucleotide-binding motif A (P-loop)  
 F;442-447/Region: nucleotide-binding motif B  
 F;702/Active site: Ser #status predicted

Query Match 1.0%; Score 8; DB 1; Length 795;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 ALDKLLER 473  
 |||||  
 Db 127 ALDKLLER 134

## RESULT 22

T00818  
 hypothetical protein At2g41620 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein T32G6.14  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 23-Mar-2001  
 C;Accession: T00818; A84844  
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
 submitted to the EMBL Data Library, November 1997  
 A;Description: Arabidopsis thaliana chromosome II BAC T32G6 genomic sequence.  
 A;Reference number: Z14163

A;Accession: T00818  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-825 <RQ>  
 A;Cross-references: EMBL:AC002510; NID:g2618683; PID:g2618698  
 A;Experimental source: cultivar Columbia  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Varakan, S.E.; Umayam, L.; Tallon, L.;  
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: A84844  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-825 <STO>  
 A;Cross-references: GB:AE002093; NID:g2618698; PIDN:AAB84345.1; GSPDB:GN00139  
 C;Genetics:

A;Gene: At2g41620; T32G6.14  
 A;Map position: 2  
 A;Introns: 43/3; 74/2; 96/3; 116/3; 146/3; 234/3; 266/3; 311/3; 329/3; 368/3; 669/3; 725  
 A;Note: T32G6.14  
 C;Superfamily: Arabidopsis thaliana hypothetical protein At2g41620

Query Match 1.0%; Score 8; DB 2; Length 825;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 REGINAEQ 48  
 |||||  
 Db 78 REGINAEQ 85

## RESULT 23

C75313  
 Probable proteinase - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans  
 C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C;Accession: C75313  
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: C75313  
 A;Status: preliminary  
 A;Molecule type: DNA

A;Residues: 1-840 <WHI>  
 A;Cross-references: GB:AE002047; GB:AE000513; NID:g6459915; PIDN:AAF11680.1; PID:g645992  
 F;442-447/Region: nucleotide-binding motif B  
 F;702/Active site: Ser #status predicted

A;Gene: DR2130  
 A;Map position: 1

Query Match 1.0%; Score 8; DB 2; Length 840;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 DGLVFDPA 294  
 |||||  
 Db 332 DGLVFDPA 339

## RESULT 24

A64714  
 helicase - Helicobacter pylori (strain 26695)

C;Species: Helicobacter pylori  
 C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
 C;Accession: A64714  
 R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney,  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: A64714  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-945 <TOM>  
 A;Cross-references: GB:AE000653; GB:AE000511; NID:g2314733; PIDN:AAD08593.1; PID:g231473

Query Match 1.0%; Score 8; DB 2; Length 945;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 ERLGKENS 506  
 |||||  
 Db 657 ERLGKENS 664

## RESULT 25

B70748  
 Probable smc protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C;Accession: B70748  
 R;Cole, S.T.; Davies, R.; Brocks, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70748  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA

A;Residues: 1-1289 <COL>  
 A;Cross-references: GB:Z74697; GB:AL123456; NID:g3261602; PIDN:CAA98982.1; PID:gl405958  
 A;Experimental source: strain H37RV

C;Genetics:  
 A;Gene: smc  
 C;Superfamily: chromosome segregation protein SMCI

Query Match 1.0%; Score 8; DB 2; Length 1289;  
 Best Local Similarity 100.0%; Pred. No. 60;



Query Match 1.0%; Score 8; DB 2; Length 292;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157  
|||||  
Db 150 NDGAVALA 157

RESULT 17  
T27791  
hypothetical protein ZK228.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C;Accession: T27791  
R;Basham, V.  
submitted to the EMBL Data Library, November 1996  
A;Reference number: Z20419  
A;Accession: T27791  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-294 <WIL>  
A;Cross-references: EMBL:Z82086; PIDN:CAB04996.1; GSPDB:GN00023; CESP:ZK228.4  
C;Genetics:  
A;Gene: CESP:ZK228.4  
A;Map position: 5  
A;Introns: 38/1; 99/1; 143/3; 181/3; 250/2  
C;Superfamily: Caenorhabditis elegans hypothetical protein F31F7.1

Query Match 1.0%; Score 8; DB 2; Length 294;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRV 18  
|||||  
Db 262 TVKENNRV 269

RESULT 18  
S05355  
hypothetical protein (clone AAC11) - slime mold (Dictyostelium discoideum) (fragment)  
C;Species: Dictyostelium discoideum  
C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 29-Oct-1999  
C;Accession: S05355  
R;Shaw, D.R.; Richter, H.; Giorda, R.; Ohmachi, T.; Ennis, H.L.  
Mol. Gen. Genet. 218, 453-459, 1989  
A;Title: Nucleotide sequences of Dictyostelium discoideum developmentally regulated cDNA  
A;Reference number: S05355; MUID:90066348; PMID:2511421  
A;Accession: S05355  
A;Molecule type: mRNA  
A;Residues: 1-448 <SHA>  
A;Cross-references: EMBL:X16522; NID:g7173; PIDN:CAA34529.1; PID:g7174

Query Match 1.0%; Score 8; DB 2; Length 448;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 NTSNNSNT 257  
|||||  
Db 378 NTSNNSNT 385

RESULT 19  
F82203  
probable multidrug resistance protein VCI409 [imported] - Vibrio cholerae (strain Ni6961)  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: F82203  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermlolaeva, M.D.; Vanathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A;Reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: F82203  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-484 <HEI>  
A;Cross-references: GB:AE004219; GB:AE003852; NID:g9555899; PIDN:AAF94566.1; GSPDB:GN0012  
A;Experimental source: serogroup O1; strain Ni6961; biotype El Tor  
C;Genetics:  
A;Gene: VCI409  
A;Map position: 1

Query Match 1.0%; Score 8; DB 2; Length 484;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 SELAAAEA 210  
|||||  
Db 169 SELAAAEA 176

RESULT 20  
B89957  
hypothetical protein ptAA [imported] - *Staphylococcus aureus* (strain N315)  
C;Species: *Staphylococcus aureus*  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: B89957  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi,  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; H  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: B89957  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-488 <KUR>  
A;Cross-references: GB:BA000018; PID:g13701521; PIDN:BAB42815.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: ptAA  
C;Superfamily: phosphotransferase system glucose-specific enzyme II, factor II; phosphot

Query Match 1.0%; Score 8; DB 2; Length 488;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 NDGAVALA 157  
|||||  
Db 72 NDGAVALA 79

RESULT 21  
S73830  
endopeptidase La (EC 3.4.21.53) - *Mycoplasma pneumoniae* (strain ATCC 29342)  
N;Alternate names: ATP-dependent proteinase lon; ATP-dependent serine proteinase La; hylp  
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)  
C;Species: *Mycoplasma pneumoniae*  
A;Variety: ATCC 29342  
C;Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 02-Jun-2003  
C;Accession: S73830  
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkle, E.; Li, B.C.; Herrmann, R.  
Nucleic Acids Res. 24, 4420-4449, 1996  
A;Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*.  
A;Reference number: S73327; MUID:97105885; PMID:8948633  
A;Accession: S73830  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-795 <HIM>  
A;Cross-references: EMBL:AF000050; GB:U00089; NID:g1674197; PIDN:AAB96152.1; PID:g1674196  
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C;Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP

Cell 72, 247-260, 1993  
 A:Title: Molecular cloning and functional analysis of Drosophila TAF110 reveal properties  
 A:Reference number: A45183; MUID:93145326; PMID:7678780  
 A:Accession: A45183  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-921 <HOE>  
 A:Experimental source: embryo  
 A>Note: sequence extracted from NCBI backbone (NCBIP:123832)  
 C:Genetics:  
 A:Gene: FlyBase:Taf110  
 A:Cross-references: FlyBase:FBgn0010280

Query Match 1.1%; Score 9; DB 2; Length 921;  
 Best Local Similarity 100.0%; Pred. No. 4.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TNTSNNST 257  
 DB 178 TNTSNNST 186  
 |||||

RESULT 13  
 E72853  
 A:Title: Autographa californica nuclear polyhedrosis virus  
 C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV  
 A>Note: dsDNA virus  
 C:Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 12-Nov-1999  
 C:Accession: E72853  
 R:AYres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.  
 Virology 202, 586-605, 1994  
 A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.  
 A:Reference number: A72850; MUID:94303173; PMID:8030224  
 A:Accession: E72853  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-71 <AY>  
 A:Cross-references: GB:L22858; NID:g510708; PIDN:AAA66659.1; PID:g559098  
 C:Genetics:  
 A:Gene: AcOrf-29

Query Match 1.0%; Score 8; DB 2; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 BEINRQKQ 137  
 DB 20 BEINRQKQ 27  
 |||||

RESULT 14  
 S58921  
 tropomyosin isoform TMI - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 20-Jun-2000  
 C:Accession: S58921; S62127; S58918  
 R:Kagawa, H.; Sugimoto, K.; Matsumoto, H.; Inoue, T.; Inadzu, H.; Takuwa, K.; Sakube, Y.  
 J. Mol. Biol. 251, 603-613, 1995  
 A:Title: Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenorhabditis elegans  
 A:Reference number: S58918; MUID:95395840; PMID:7666414  
 A:Accession: S58921  
 A:Molecule type: mRNA  
 A:Residues: 1-276, 'NLP', 280-284 <KAF>  
 A:Cross-references: EMBL:D38540; NID:g1208412; PIDN:BAA07543.1; PID:g1208413  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-276, 'NLP', 280-284 <KAF>  
 A:Cross-references: EMBL:D38539; NID:g871835; PIDN:BAA07540.1; PID:g1208409  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994  
 A:Accession: S58918  
 A:Molecule type: protein  
 A:Residues: 11-49; 171-211 <KAW>

C:Genetics:  
 A:Gene: tmy-1  
 A:Introns: 15/3; 61/3; 80/3; 125/2; 164/3; 188/2; 234/3; 258/1  
 C:Superfamily: tropomyosin  
 C:Keywords: alternative splicing

Query Match 1.0%; Score 8; DB 2; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAQLKEA 734  
 DB 144 VEAQLKEA 151  
 |||||

RESULT 15  
 S58922  
 tropomyosin isoform TMI - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 20-Jun-2000  
 C:Accession: S58922; S62128; S58918  
 R:Kagawa, H.; Sugimoto, K.; Matsumoto, H.; Inoue, T.; Inadzu, H.; Takuwa, K.; Sakube, Y.  
 J. Mol. Biol. 251, 603-613, 1995  
 A:Title: Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenorhabditis elegans  
 A:Reference number: S58918; MUID:95395840; PMID:7666414  
 A:Accession: S58922  
 A:Molecule type: mRNA  
 A:Residues: 1-284 <KAG>  
 A:Cross-references: EMBL:D38541; NID:g871837; PIDN:BAA07544.1; PID:g1208414  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-284 <KAW>  
 A:Cross-references: EMBL:D38539; NID:g871835; PIDN:BAA07541.1; PID:g1208410  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994  
 A:Accession: S58918  
 A:Molecule type: protein  
 A:Residues: 11-49; 171-211 <KAF>  
 C:Genetics:  
 A:Gene: tmy-1  
 A:Introns: 15/3; 61/3; 80/3; 125/2; 164/3; 188/2; 234/3; 258/1  
 C:Superfamily: tropomyosin  
 C:Keywords: alternative splicing

Query Match 1.0%; Score 8; DB 2; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 VEAQLKEA 734  
 DB 144 VEAQLKEA 151  
 |||||

RESULT 16  
 T49361  
 probable transport vesicle fusion protein SEC17 [imported] - Neurospora crassa  
 N:Alternate names: protein B1D1.150  
 C:Species: Neurospora crassa  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 A:Accession: T49361  
 R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, R.  
 submitted to the Protein Sequence Database, May 2000  
 A:Reference number: Z25022  
 A:Accession: T49361  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-292 <SCH>  
 A:Cross-references: EMBL:AL355927; GSPDB:GN00116; NCSP:B1D1.150  
 A:Experimental source: BAC clone B1D1; strain OR74A  
 C:Genetics:  
 A:Gene: NCSP:B1D1.150  
 A:Map position: 6  
 A:Introns: 11/3; 53/1; 155/2

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A;Reference number: A95000; MUID:21357203; PMID:11463916  
A;Accession: H95115

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1039 <KUR>  
A;Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:SP4  
A;Experimental source: strain TIGR4  
C;Genetics:  
A;Gene: SP1004

Query Match 4.0%; Score 32; DB 2; Length 1039;  
Best Local Similarity 100.0%; Pred. No. 2.3e-23;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 77  
|||||  
Db 67 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 98  
|||||

RESULT 8  
D97985

hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C;Accession: D97985  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: D97985  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1039 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174  
C;Genetics:  
A;Gene: phtE

Query Match 4.0%; Score 32; DB 2; Length 1039;  
Best Local Similarity 100.0%; Pred. No. 2.3e-23;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 77  
|||||  
Db 67 AEOIVIKITDQGVVTSHTGHDHYHYNGKVPYDA 98  
|||||

RESULT 9  
T46758

hypothetical 92.4K protein - Streptococcus agalactiae  
C;Species: Streptococcus agalactiae  
C;Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000

C;Accession: T46758  
R;Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heynemann, J.; Schnitzler, N.; Lued  
Infect. Immun. 67, 871-878, 1999  
A;Title: Lmb, a protein with similarities to the Irai adhesin family, mediates attachmen  
A;Reference number: 224091; MUID:99115569; PMID:9916102  
A;Accession: T46758  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-822 <SPE>  
A;Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AAD13797.1; PID:g4249624  
A;Experimental source: strain R268  
C;Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 2.1%; Score 17; DB 2; Length 822;  
Best Local Similarity 100.0%; Pred. No. 3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISEELM 85  
|||||  
Db 92 YNGKVPYDAIISEELM 108  
|||||

RESULT 10  
F97985

hypothetical protein phtE-truncation [imported] - Streptococcus pneumoniae (strain R6)  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C;Accession: F97985  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: F97985  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-182 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:g15458517; GSPDB:GN00174  
C;Genetics:  
A;Gene: phtE-truncation

Query Match 1.3%; Score 10; DB 2; Length 182;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 HGDHYHYIPK 197  
|||||  
Db 35 HGDHYHYIPK 44  
|||||

RESULT 11  
A48184

transcription initiation factor IID 110K chain - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 16-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999

C;Accession: A48184  
R;Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.  
Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993  
A;Title: The Drosophila 110-kDa transcription factor TFIIID subunit directly interacts wit  
A;Reference number: A48184; MUID:93317591; PMID:8327460  
A;Accession: A48184  
A;Status: preliminary  
A;Molecule type: mRNA; protein  
A;Residues: 1-921 <KOK>  
A;Cross-references: NID:g398432; PID:g398433  
A;Experimental source: embryo nuclear extract  
A;Note: sequence extracted from NCBI backbone (NCBIN:134863, NCBIIP:134864)  
C;Genetics:  
A;Gene: FlyBase:Taf110  
A;Cross-references: FlyBase:FBgn0010280  
C;Keywords: transcription initiation

Query Match 1.1%; Score 9; DB 2; Length 921;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 TMTSNNST 257  
|||||  
Db 178 TMTSNNST 186  
|||||

RESULT 12  
A45183

TBP-associated factor TFIIID - fruit fly (Drosophila sp.)  
C;Species: Drosophila sp.  
C;Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Feb-1997  
C;Accession: A45183  
R;Hoey, T.; Weinzierl, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.



on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappple, nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A:Authors: Lotus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: C95136

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-802 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:g14972655; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1175

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 87.3%; Score 695; DB 2; Length 802;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60  
Db 7 SYELGLYQARTVKNRNVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 66  
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNXYKLKDEDIVNEVKGYYVVKDGYVYVLKDA 120  
Db 67 SHGDHYHYNGKVPYDAIISEELMKDPNXYKLKDEDIVNEVKGYYVVKDGYVYVLKDA 126  
QY 121 AHADNVRTKEINRQKQHSQHREGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 180  
Db 127 AHADNVRTKEINRQKQHSQHREGGTPRNDGAVALARSQGYTTDDGYIFNASDIIEDT 186  
QY 181 GDVIVPHGDHYHYVYPKNEISASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240  
Db 187 GDVIVPHGDHYHYVYPKNEISASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 246  
QY 241 VSNPQTNTNTSNNTNSQASQNDIDSLKQYKLPLOSRHVESDGLVDPQAITSRT 300  
Db 247 VSNPQTNTNTSNNTNSQASQNDIDSLKQYKLPLOSRHVESDGLVDPQAITSRT 306  
QY 301 ARGVAVPHGDHYHYPIYQSOMSELEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360  
Db 307 ARGVAVPHGDHYHYPIYQSOMSELEERARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 366  
QY 361 PQAPNLKIDNSLSVLSQYKVGEGYVFEKGISRVVFAKLPSETVKNLESKLSQES 420  
Db 367 PQAPNLKIDNSLSVLSQYKVGEGYVFEKGISRVVFAKLPSETVKNLESKLSQES 426  
QY 421 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 480  
Db 427 VSHLTAKENVAPRDQEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTN 486  
QY 481 KEKLVDDLLAFAPITHPERLGPNSQIETDEVRIAQLADKYTTSDGYIFDEHDIISD 540  
Db 487 KEKLVDDLLAFAPITHPERLGPNSQIETDEVRIAQLADKYTTSDGYIFDEHDIISD 546  
QY 541 EGAAYVTPHGHSHWICKDSLSDEKVAQAQYTKEGILPSPDADYKANTPGDSAAIY 600  
Db 547 EGAAYVTPHGHSHWICKDSLSDEKVAQAQYTKEGILPSPDADYKANTPGDSAAIY 606  
QY 601 NRKVGKRIIPLVRLPYMVEHTEVKNGLNLIIPHKDHYHNIKFAMFDDHTYKAPNGYTTLED 660  
Db 607 NRKVGKRIIPLVRLPYMVEHTEVKNGLNLIIPHKDHYHNIKFAMFDDHTYKAPNGYTTLED 666  
QY 661 LFATIKYVVEHPDRPHNSDGNWASHVILGKKDHSDEPNKNFKADEPVEETPAEPEVP 720  
Db 667 LFATIKYVVEHPDRPHNSDGNWASHVILGKKDHSDEPNKNFKADEPVEETPAEPEVP 726  
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLIITLQIMDNNSIMAEAEKLLA 780  
Db 727 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLIITLQIMDNNSIMAEAEKLLA 786  
QY 781 LLKGSNPSSVSKKIN 796

Db 787 LLKGSNPSSVSKKIN 802

RESULT 2

E98004

hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: E98004

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Ee

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, B.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: E98004

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-828 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:g15458683; GSPDB:GN00174

C:Genetics:

A:Gene: phtA

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 77.4%; Score 616; DB 2; Length 828;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 716; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 SEELMKDPNXYKLKDEDIVNEVKGYYVVKDGYVYVLKDAHADNVRTKEINRQKQEH 139  
Db 112 SEELMKDPNXYKLKDEDIVNEVKGYYVVKDGYVYVLKDAHADNVRTKEINRQKQEH 171  
QY 140 SQHREGTTPRNDGAVALARSQGYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNE 199  
Db 172 SQHREGTTPRNDGAVALARSQGYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNE 231  
QY 200 LSASELAAAEAFISGRGNLSNRTYRRQNSDNTSRTNWVPSVSNPGTTNTNNSNTNS 259  
Db 232 LSASELAAAEAFISGRGNLSNRTYRRQNSDNTSRTNWVPSVSNPGTTNTNNSNTNS 291  
QY 260 QASQNDIDSLKQYKLPLOSRHVESDGLVDPQAITSRTARGVAVPHGDHYHYIPYSQ 319  
Db 292 QASQNDIDSLKQYKLPLOSRHVESDGLVDPQAITSRTARGVAVPHGDHYHYIPYSQ 351  
QY 320 MSELERARIIPLYRSNHWVPSRPEQSPQPTPEPSGPAPNLKIDNSLSVLSQ 379  
Db 352 MSELERARIIPLYRSNHWVPSRPEQSPQPTPEPSGPAPNLKIDNSLSVLSQ 411  
QY 380 VRKVGEGYVFEKGISRYVFAKLPSETVKNLESKLSQESVSHLTAKENVAPRDQEF 439  
Db 412 VRKVGEGYVFEKGISRYVFAKLPSETVKNLESKLSQESVSHLTAKENVAPRDQEF 471  
QY 440 YDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPITHP 499  
Db 472 YDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPITHP 531  
QY 500 RLKGPSQIETDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHGHSHWICKD 559  
Db 532 RLKGPSQIETDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHGHSHWICKD 591  
QY 560 SLSDKEKVAQAQYTKEGILPSPDADYKANTPGDSAAAIYNRVKGEKRIPLVRLPYMVE 619  
Db 592 SLSDKEKVAQAQYTKEGILPSPDADYKANTPGDSAAAIYNRVKGEKRIPLVRLPYMVE 651  
QY 620 HTVEVKNGLNLIIPHKDHYHNIKFAMFDDHTYKAPNGYTTLEDLFTIKYVVEHDEPHSN 679  
Db 652 HTVEVKNGLNLIIPHKDHYHNIKFAMFDDHTYKAPNGYTTLEDLFTIKYVVEHDEPHSN 711  
QY 680 DGMGNASEHVLGKKDHSDEPNKNFKADEPVEETPAEPEVPQVETEKVEAQLKEAEVLLA 739  
Db 712 DGMGNASEHVLGKKDHSDEPNKNFKADEPVEETPAEPEVPQVETEKVEAQLKEAEVLLA 771

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:20:59 ; Search time 27 Seconds  
(without alignments)  
2835.869 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SVELGLYQARTVKENRVS.....KLALLKGNPSSVSKEKIN 796

Scoring table:

OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

PIR\_78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	695	87.3	802	C95136	conserved domain p
2	616	77.4	828	E98004	hypothetical prote
3	107	13.4	855	D98004	histidine Motif-Co
4	60	7.5	819	B95136	conserved domain p
5	60	7.5	839	G95115	conserved hypothet
6	60	7.5	853	C97985	hypothetical prote
7	32	4.0	1039	H95115	conserved hypothet
8	32	4.0	1039	D97985	hypothetical prote
9	17	2.1	822	T46758	hypothetical 92.4K
10	10	1.3	182	F97985	hypothetical prote
11	9	1.1	921	A48184	transcription init
12	9	1.1	921	A45183	TBP-associated fac
13	8	1.0	71	E72853	AcOrf-29 protein -
14	8	1.0	284	S58921	tropomyosin isofor
15	8	1.0	284	S58922	tropomyosin isofor
16	8	1.0	292	T49361	probable transport
17	8	1.0	294	T27791	hypothetical prote
18	8	1.0	448	S05355	hypothetical prote
19	8	1.0	484	F82203	probable multidrug
20	8	1.0	488	B89957	hypothetical prote
21	8	1.0	795	S73830	endopeptidase Ia (
22	8	1.0	825	T00818	hypothetical prote
23	8	1.0	840	C75313	probable proteinas
24	8	1.0	945	A64714	helicase - Helicob
25	8	1.0	1289	B70748	probable smc prote
26	8	1.0	1659	H97926	hypothetical prote
27	8	1.0	1659	G95057	endo-beta-N-acetyl
28	7	0.9	44	S55073	env polyprotein -
29	7	0.9	49	S55072	env polyprotein -

## ALIGNMENTS

## RESULT 1

C95136 conserved domain protein SP1175 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: C95136  
R:Tetreltin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

30	7	0.9	49	2	D70091	hypothetical prote
31	7	0.9	68	2	D87657	hypothetical prote
32	7	0.9	70	2	A75134	hypothetical prote
33	7	0.9	90	2	C90404	conserved hypothet
34	7	0.9	93	2	C75396	hypothetical prote
35	7	0.9	94	2	E75556	hypothetical prote
36	7	0.9	100	1	BVNYBA	chaperonin groES -
37	7	0.9	100	1	BVNY7B	chaperonin groES -
38	7	0.9	101	2	T38824	thioltransferase -
39	7	0.9	101	2	E82874	ribosomal protein
40	7	0.9	106	2	I64171	hypothetical prote
41	7	0.9	110	2	D87498	HesB/yadR/yfjH fam
42	7	0.9	110	2	C86891	hypothetical prote
43	7	0.9	117	2	AG3598	transcription regu
44	7	0.9	121	2	T44473	conserved hypothet
45	7	0.9	124	1	NRPRH	pancreatic ribonuc
46	7	0.9	128	2	H72500	hypothetical prote
47	7	0.9	129	2	T36220	mutr-like protein
48	7	0.9	130	2	S76934	hypothetical prote
49	7	0.9	131	2	T35461	mutr-like protein
50	7	0.9	136	2	B65240	hypothetical prote
51	7	0.9	136	2	D64964	hypothetical prote
52	7	0.9	136	2	T19240	hypothetical prote
53	7	0.9	139	2	B81286	probable membrane
54	7	0.9	143	2	A71717	hypothetical prote
55	7	0.9	145	2	AF1219	Salmonella enteric
56	7	0.9	147	2	D83261	probable transcrip
57	7	0.9	148	2	A86079	PTS system, fructo
58	7	0.9	148	2	B91232	PTS system, fructo
59	7	0.9	148	2	D48649	frva protein - Esc
60	7	0.9	154	2	S39873	hypothetical prote
61	7	0.9	157	2	AC2947	ribonuclease H [im
62	7	0.9	157	2	G98335	ribonuclease H PAI
63	7	0.9	157	2	AF3240	conserved hypothet
64	7	0.9	158	1	D63256	hypothetical prote
65	7	0.9	158	2	A89943	transcription elon
66	7	0.9	165	2	C83415	hypothetical prote
67	7	0.9	166	2	A86450	probable glycine c
68	7	0.9	178	2	C81235	transcription anti
69	7	0.9	179	2	H71920	adenine phosphorib
70	7	0.9	181	2	A70178	conserved hypothet
71	7	0.9	185	2	F75487	v-type ATP synthas
72	7	0.9	191	2	T35792	beta-glucosidase -
73	7	0.9	197	1	S01249	cytochrome c-L pre
74	7	0.9	201	2	S06147	GTP-binding protei
75	7	0.9	203	2	F90658	hypothetical prote
76	7	0.9	203	2	F85509	hypothetical prote
77	7	0.9	210	2	S62443	hypothetical prote
78	7	0.9	211	2	AH0899	diadenosine 5', 5'
79	7	0.9	211	2	T01627	conserved hypothet
80	7	0.9	213	2	F83144	probable ATP bindi
81	7	0.9	219	2	B75533	conserved hypothet
82	7	0.9	226	2	D65105	hypothetical prote
83	7	0.9	226	2	A98133	hypothetical prote
84	7	0.9	226	2	D85978	hypothetical prote
85	7	0.9	226	2	G81894	probable integral
86	7	0.9	232	2	D81125	hypothetical prote
87	7	0.9	232	2	C83744	two-component resp
88	7	0.9	237	2	G87286	conserved hypothet
89	7	0.9	238	2	I67638	proteasome activat
90	7	0.9	240	2	D75203	hypothetical prote

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 90  
 |||||  
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 92

## RESULT 24

AAB12763  
 ID AAB12763 standard; protein; 811 AA.  
 XX  
 AC AAB12763;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae strain RX1 BVH-11 protein antigen.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200039299-A2.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 20-DEC-1999; 99WO-CA001218.  
 XX  
 PR 23-DEC-1998; 98US-0113800P.  
 XX  
 PA (BIOC-) BIOCHEM PHARMA INC.  
 XX  
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX WPI; 2000-452397/39.  
 DR

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otis media, bacteraemia and/or pneumonia.  
 XX  
 PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
 CC antigen, from the present invention  
 XX

SQ Sequence 811 AA;

Query Match 7.5%; Score 60; DB 3; Length 811;  
 Best Local Similarity 100.0%; Pred. No. 2e-48;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 90  
 |||||  
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 92

## RESULT 25

AAB12758  
 ID AAB12758 standard; protein; 816 AA.  
 XX  
 AC AAB12758;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae strain WU2 BVH-11-2 protein antigen.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200039299-A2.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 20-DEC-1999; 99WO-CA001218.  
 XX  
 PR 23-DEC-1998; 98US-0113800P.  
 XX  
 PA (BIOC-) BIOCHEM PHARMA INC.  
 XX

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX WPI; 2000-452397/39.  
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otis media, bacteraemia and/or pneumonia.  
 XX  
 PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11-2  
 CC protein antigen, from the present invention  
 XX

SQ Sequence 816 AA;

Query Match 7.5%; Score 60; DB 3; Length 816;  
 Best Local Similarity 100.0%; Pred. No. 2e-48;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 90  
 |||||  
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHNGKVPYDAIISELLMKDPNY 92

Search completed: October 1, 2004, 07:32:24  
 Job time : 84 secs

PN WO200039299-A2.  
XX  
PD 06-JUL-2000.  
XX  
PF 20-DEC-1999; 99WO-CA001218.  
XX  
PR 23-DEC-1998; 98US-0113800P.  
XX  
PA (BIOC-) BIOCHEM PHARMA INC.  
XX  
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
XX  
DR WPI; 2000-452397/39.  
XX  
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia.  
XX  
PS Disclosure; Fig 12; 106pp; English.  
XX  
CC The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
CC may be used for the recombinant production of the proteins they encode.  
CC The protein antigens may then be used as vaccines for the prevention and  
CC treatment of Streptococcal infections in mammals (especially humans)  
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
CC antigen, from the present invention  
XX  
XX Sequence 811 AA;  
SQ  
Query Match 7.5%; Score 60; DB 3; Length 811;  
Best Local Similarity 100.0%; Pred. No. 2e-48;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90  
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92  
RESULT 22  
AAB12762  
ID AAB12762 standard; protein; 811 AA.  
AC AAB12762;  
XX  
XX 21-NOV-2000 (first entry)  
XX  
XX Streptococcus pneumoniae strain A66 BVH-11 protein antigen.  
XX  
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX WO200039299-A2.  
XX  
XX 06-JUL-2000.  
XX  
XX 20-DEC-1999; 99WO-CA001218.  
XX  
XX 23-DEC-1998; 98US-0113800P.  
XX  
XX (BIOC-) BIOCHEM PHARMA INC.  
XX  
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
XX  
XX WPI; 2000-452397/39.  
XX  
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia.  
XX  
PS Disclosure; Fig 12; 106pp; English.  
XX  
XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
CC may be used for the recombinant production of the proteins they encode.  
CC The protein antigens may then be used as vaccines for the prevention and  
CC treatment of Streptococcal infections in mammals (especially humans)  
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
CC antigen, from the present invention  
XX  
XX Sequence 811 AA;  
SQ  
Query Match 7.5%; Score 60; DB 3; Length 811;  
Best Local Similarity 100.0%; Pred. No. 2e-48;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90  
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92  
RESULT 22  
AAB12762  
ID AAB12762 standard; protein; 811 AA.  
AC AAB12762;  
XX  
XX 21-NOV-2000 (first entry)  
XX  
XX Streptococcus pneumoniae strain A66 BVH-11 protein antigen.  
XX  
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX WO200039299-A2.  
XX  
XX 06-JUL-2000.  
XX  
XX 20-DEC-1999; 99WO-CA001218.  
XX  
XX 23-DEC-1998; 98US-0113800P.  
XX  
XX (BIOC-) BIOCHEM PHARMA INC.  
XX  
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
XX  
XX WPI; 2000-452397/39.  
XX  
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia.  
XX

PS Disclosure; Fig 12; 106pp; English.  
XX  
CC The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
CC may be used for the recombinant production of the proteins they encode.  
CC The protein antigens may then be used as vaccines for the prevention and  
CC treatment of Streptococcal infections in mammals (especially humans)  
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
CC antigen, from the present invention  
XX  
XX Sequence 811 AA;  
SQ  
Query Match 7.5%; Score 60; DB 3; Length 811;  
Best Local Similarity 100.0%; Pred. No. 2e-48;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90  
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92  
RESULT 23  
AAB12760  
ID AAB12760 standard; protein; 811 AA.  
XX  
XX AAB12760;  
XX  
XX 21-NOV-2000 (first entry)  
XX  
XX Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.  
XX  
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX WO200039299-A2.  
XX  
XX 06-JUL-2000.  
XX  
XX 20-DEC-1999; 99WO-CA001218.  
XX  
XX 23-DEC-1998; 98US-0113800P.  
XX  
XX (BIOC-) BIOCHEM PHARMA INC.  
XX  
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
XX  
XX WPI; 2000-452397/39.  
XX  
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia.  
XX  
PS Disclosure; Fig 12; 106pp; English.  
XX  
XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
CC may be used for the recombinant production of the proteins they encode.  
CC The protein antigens may then be used as vaccines for the prevention and  
CC treatment of Streptococcal infections in mammals (especially humans)  
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
CC antigen, from the present invention  
XX  
XX Sequence 811 AA;  
SQ  
Query Match 7.5%; Score 60; DB 3; Length 811;  
Best Local Similarity 100.0%; Pred. No. 2e-48;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 31 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 90  
DB 33 ENLTDEVSKREGINAEQIVIKITQGYVTSHTGDHYHYNGKVPYDAIISELLMKDPNY 92  
RESULT 23  
AAB12760  
ID AAB12760 standard; protein; 811 AA.  
XX  
XX AAB12760;  
XX  
XX 21-NOV-2000 (first entry)  
XX  
XX Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.  
XX  
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal.  
XX  
OS Streptococcus pneumoniae.  
XX  
XX WO200039299-A2.  
XX  
XX 06-JUL-2000.  
XX  
XX 20-DEC-1999; 99WO-CA001218.  
XX  
XX 23-DEC-1998; 98US-0113800P.  
XX  
XX (BIOC-) BIOCHEM PHARMA INC.  
XX  
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
XX  
XX WPI; 2000-452397/39.  
XX  
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia.  
XX  
PS Disclosure; Fig 12; 106pp; English.  
XX  
XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
CC may be used for the recombinant production of the proteins they encode.  
CC The protein antigens may then be used as vaccines for the prevention and  
CC treatment of Streptococcal infections in mammals (especially humans)  
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
CC antigen, from the present invention  
XX  
XX Sequence 811 AA;  
SQ  
Query Match 7.5%; Score 60; DB 3; Length 811;  
Best Local Similarity 100.0%; Pred. No. 2e-48;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



CC represented in the printed specification, but is based on sequence  
 CC information supplied by the European Patent Office  
 XX  
 SQ Sequence 840 AA;

Query Match 10.1%; Score 80; DB 7; Length 840;  
 Best Local Similarity 100.0%; Pred. No. 1e-67;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70  
 |||||  
 Db 31 TVKENRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 90  
 |||||  
 QY 71 GKVPYDAIISELLMKDPNY 90  
 |||||  
 Db 91 GKVPYDAIISELLMKDPNY 110

RESULT 19  
 AAB12764  
 ID AAB12764 standard; protein; 805 AA.  
 XX  
 AC AAB12764;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae strain JNR7/87 BVH-11 protein antigen.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200039299-A2.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 20-DEC-1999; 99WO-CA001218.  
 XX  
 PR 23-DEC-1998; 98US-0113800P.  
 XX  
 PA (BIOC-) BIOCHEM PHARMA INC.  
 XX  
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX  
 DR WPI; 2000-452397/39.

XX Streptococcus pneumoniae.  
 XX  
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otitis media, bacteraemia and/or pneumonia.  
 XX  
 PS Disclosure; Fig 12; 106pp; English.  
 XX  
 CC The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
 CC antigen, from the present invention  
 XX  
 SQ Sequence 805 AA;

Query Match 7.5%; Score 60; DB 3; Length 805;  
 Best Local Similarity 100.0%; Pred. No. 2e-48;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 90  
 |||||  
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 92  
 |||||

RESULT 20  
 AAB12765  
 ID AAB12765 standard; protein; 807 AA.  
 XX  
 AC AAB12765;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae strain SP63 BVH-11 protein antigen.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200039299-A2.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 20-DEC-1999; 99WO-CA001218.  
 XX  
 PR 23-DEC-1998; 98US-0113800P.  
 XX  
 PA (BIOC-) BIOCHEM PHARMA INC.  
 XX  
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX  
 DR WPI; 2000-452397/39.  
 XX  
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otitis media, bacteraemia and/or pneumonia.  
 XX  
 PS Disclosure; Fig 12; 106pp; English.  
 XX  
 CC The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
 CC antigen, from the present invention  
 XX  
 SQ Sequence 807 AA;

Query Match 7.5%; Score 60; DB 3; Length 807;  
 Best Local Similarity 100.0%; Pred. No. 2e-48;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 90  
 |||||  
 Db 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNKVPYDAIISELLMKDPNY 92  
 |||||

RESULT 21  
 AAB12761  
 ID AAB12761 standard; protein; 811 AA.  
 XX  
 AC AAB12761;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX

DE Streptococcus pneumoniae strain WU2 BVH-11 protein antigen.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal.  
 XX  
 OS Streptococcus pneumoniae.  
 XX

XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
CC bactericidal activity. The nucleic acids, encoding the protein antigens, have  
CC may be used for the recombinant production of the proteins they encode.  
CC The protein antigens may then be used as vaccines for the prevention and  
CC treatment of Streptococcal infections in mammals (especially humans)  
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
CC pneumonia. The present sequence represents the S. pneumoniae BVH-11  
CC protein antigen  
XX  
XX  
SQ Sequence 840 AA;  
Query Match 10.1%; Score 80; DB 3; Length 840;  
Best Local Similarity 100.0%; Pred. No. 1e-67;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRSHGHYHYN 70  
DB 31 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRSHGHYHYN 90  
QY 71 GKVPYDAIISBELLMKDPNY 90  
DB 91 GKVPYDAIISBELLMKDPNY 110  
RESULT 17  
AAU75933  
ID AAU75933 standard; protein; 840 AA.  
XX  
AC AAU75933;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Streptococcus pneumoniae BVH-11 protein.  
XX  
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;  
KW streptococcal bacterial infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200198334-A2.  
XX  
PD 27-DEC-2001.  
XX  
PF 19-JUN-2001; 2001WO-CA000909.  
XX  
PR 20-JUN-2000; 2000US-0212683P.  
XX  
PA (SHIR-) SHIRE BIOCHEM INC.  
XX  
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;  
XX  
DR WPI; 2002-122272/16.  
DR N-PSDB; ABK15103.  
XX  
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing  
PT polypeptides, useful as vaccine components for treating or preventing  
PT streptococcal infections such as otitis media, meningitis, and  
PT bacteraemia.  
XX  
XX Example 1; Fig 7; 113pp; English.  
XX  
CC The invention describes an isolated polypeptide (I) with 70-90% identity  
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or  
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)  
CC comprising (I) is useful for therapeutic or prophylactic treatment of  
CC meningitis, otitis media, bacteraemia or pneumonia infection in an  
CC individual susceptible to these disorders. (II) is also useful for  
CC therapeutic or prophylactic treatment of any streptococcal bacterial  
CC infection (e.g., caused by Streptococcus pneumoniae, group A  
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such  
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or

CC Staphylococcus aureus) in an individual susceptible to the infection. A  
CC polynucleotide (III) encoding (I) is useful in DNA immunisation  
CC techniques. The Streptococcus polypeptides are useful in a diagnostic  
CC test for S. pneumoniae infection. (III) is useful for designing DNA  
CC probes for use in detecting the presence of Streptococcus in a biological  
CC sample suspected of containing the bacteria. The DNA probes may also be  
CC used for detecting circulating S. pneumonia nucleic acid in a sample for  
CC diagnosing streptococcal infections. This is the amino acid sequence of  
CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic  
CC peptides described in the method of the invention  
XX  
XX  
SQ Sequence 840 AA;  
Query Match 10.1%; Score 80; DB 5; Length 840;  
Best Local Similarity 100.0%; Pred. No. 1e-67;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRSHGHYHYN 70  
DB 31 TVKENRVSYIDGKQATKNTLTPDEVSKREGINAEQIVIKITDQGYVTSRSHGHYHYN 90  
QY 71 GKVPYDAIISBELLMKDPNY 90  
DB 91 GKVPYDAIISBELLMKDPNY 110  
RESULT 18  
ABM18797  
ID ABM18797 standard; protein; 840 AA.  
XX  
AC ABM18797;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
DE S. pneumoniae BVH-11 polypeptide SEQ ID NO: 8.  
XX  
KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;  
KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO2003054007-A2.  
XX  
PD 03-JUL-2003.  
XX  
PF 20-DEC-2002; 2002WO-CA002006.  
XX  
PR 20-DEC-2001; 2001US-0341252P.  
XX  
PA (SHIR-) SHIRE BIOCHEM INC.  
XX  
PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;  
XX  
DR Labbe S;  
XX  
DR WPI; 2003-569224/53.  
XX  
PT New isolated polypeptides of Streptococcus pneumoniae, useful for  
PT diagnosing, preventing or treating streptococcal infection, meningitis,  
PT otitis media, bacteraemia or pneumonia infection.  
XX  
XX Example 1; Fig 10; 79pp; English.  
XX  
CC The invention relates to a novel isolated polypeptide of Streptococcus  
CC pneumoniae. A polypeptide of the invention has antibacterial,  
CC antiinflammatory, and auditory activity, and is used as a vaccine. The  
CC polypeptide or pharmaceutical composition is useful for the prophylactic  
CC or therapeutic treatment of streptococcal infection, meningitis, otitis  
CC media, bacteraemia or pneumonia infection. The kit is useful for  
CC detecting or diagnosing streptococcal infection. The pharmaceutical  
CC composition is useful as a vaccine. The polynucleotides are useful in  
CC designing DNA probes for detecting circulating Streptococcus in a  
CC biological sample. The present sequence is used in the exemplification of  
CC the invention. Note: The sequence data for this patent is not fully



XX 21-NOV-2000 (first entry)  
 DT Streptococcus pneumoniae strain SP64 BVH-11 protein antigen.  
 XX Streptococcus pneumoniae; BVH-3; BVH-11; antigen; vaccine;  
 DE prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 XX otitis media; pneumonia; immunisation; bactericidal.  
 KW Streptococcus pneumoniae.  
 XX WO200039299-A2.  
 PN 06-JUL-2000.  
 XX 20-DEC-1999; 99WO-CA001218.  
 PF 23-DEC-1998; 98US-0113800P.  
 XX (BIOC-) BIOCHEM PHARMA INC.  
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX WPI; 2000-452397/39.  
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otitis media, bacteraemia and/or pneumonia.  
 XX Disclosure; Fig 12; 106pp; English.  
 XX The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence represents a S. pneumoniae BVH-11 protein  
 CC antigen, from the present invention  
 XX Sequence 821 AA;  
 SQ

Query Match 10.1%; Score 80; DB 3; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 1e-67;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 70  
 DB 12 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 71

QY 71 GKVPYDAIISSELLMKDPNY 90  
 DB 72 GKVPYDAIISSELLMKDPNY 91

RESULT 13  
 AAB12727  
 ID AAB12727 standard; protein; 821 AA.  
 XX AAB12727;  
 AC AAB12727;  
 DT 21-NOV-2000 (first entry)  
 DE Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.  
 XX Streptococcus pneumoniae; BVH-3; BVH-11; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal.  
 XX Streptococcus pneumoniae.  
 OS WO200039299-A2.  
 PN 21-NOV-2000 (first entry)  
 XX

PD 06-JUL-2000.  
 XX 20-DEC-1999; 99WO-CA001218.  
 XX 23-DEC-1998; 98US-0113800P.  
 XX (BIOC-) BIOCHEM PHARMA INC.  
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX WPI; 2000-452397/39.  
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otitis media, bacteraemia and/or pneumonia.  
 XX Claim 18; Fig 25; 106pp; English.  
 XX The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence represents the S. pneumoniae BVH-11M  
 CC protein antigen  
 XX Sequence 821 AA;  
 SQ

Query Match 10.1%; Score 80; DB 3; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 1e-67;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 70  
 DB 12 TVKENNRVSYIDGKQATKTNLTDPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 71

QY 71 GKVPYDAIISSELLMKDPNY 90  
 DB 72 GKVPYDAIISSELLMKDPNY 91

RESULT 14  
 AAU84026  
 ID AAU84026 standard; peptide; 821 AA.  
 XX AAU84026;  
 AC AAU84026;  
 DT 08-MAY-2002 (first entry)  
 DE Truncated variant of S. pneumoniae BVH-11, BVH-11M.  
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;  
 KW streptococcal bacterial infection; mutant; mutain.  
 XX Streptococcus pneumoniae.  
 OS Synthetic.  
 XX WO200198334-A2.  
 PN 27-DEC-2001.  
 XX 19-JUN-2001; 2001WO-CA000908.  
 PF 20-JUN-2000; 2000US-0212683P.  
 XX (SHIR-) SHIRE BIOCHEM INC.  
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;  
 XX WPI; 2002-122272/16.  
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing  
 PT

RESULT 10  
AAU84093  
ID AAU84093 standard; peptide; 690 AA.  
XX AC AAU84093;  
XX DT 08-MAY-2002 (first entry)  
XX XX  
XX DE Truncated variant of S. pneumoniae BVH-11, NEW16.  
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;  
XX KW streptococcal bacterial infection; mutant; muten.  
XX OS Streptococcus pneumoniae.  
XX XX WO200198334-A2.  
XX PN  
XX PD 27-DEC-2001.  
XX XX  
XX PF 19-JUN-2001; 2001WO-CA000908.  
XX XX 20-JUN-2000; 2000US-0212683P.  
XX PR  
XX XX (SHIR-) SHIRE BIOCHEM INC.  
XX PA  
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;  
XX DR WPI; 2002-12272/16.  
XX XX  
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing  
XX PT polypeptides, useful as vaccine components for treating or preventing  
XX PT streptococcal infections such as otitis media, meningitis, and  
XX PT bacteremia.  
XX PS  
XX PS Example 1; Page; 113pp; English.  
XX CC The invention describes an isolated polypeptide (I) with 70-90% identity  
XX CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or  
XX CC BVH-11, or chimeric sequences derived from them. A vaccine (II)  
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of  
XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an  
XX CC individual susceptible to these disorders. (II) is also useful for  
XX CC therapeutic or prophylactic treatment of any streptococcal bacterial  
XX CC infection (e.g., caused by Streptococcus pneumoniae, group A  
XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such  
XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or  
XX CC Staphylococcus aureus) in an individual susceptible to the infection. A  
XX CC polynucleotide (III) encoding (I) is useful in DNA immunisation  
XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic  
XX CC test for S. pneumoniae infection. (III) is useful for designing DNA  
XX CC probes for use in detecting the presence of Streptococcus in a biological  
XX CC sample suspected of containing the bacteria. The DNA probes may also be  
XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for  
XX CC diagnosing streptococcal infections. This sequence represents a truncate  
XX CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,  
XX CC described in the method of the invention. Note: This sequence does not  
XX CC appear in the specification but has been created according to information  
XX CC given in the invention  
XX SQ Sequence 690 AA;

Query Match 10.1%; Score 80; DB 5; Length 690;  
Best Local Similarity 100.0%; Pred. No. 8.8e-68;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 70  
DB 12 TVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90  
DB 72 GKVPYDAIISELLMKDPNY 91

RESULT 12  
AAB12766  
ID AAB12766 standard; protein; 821 AA.  
XX AC AAB12766;

RESULT 11  
ABM18826  
ID ABM18826 standard; protein; 690 AA.  
XX AC ABM18826;  
XX DT 13-OCT-2003 (first entry)  
XX XX  
XX DE S. pneumoniae variant protein NEW16.  
XX KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;  
XX KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;  
XX KW mutant; muten; New 43; BVH-11; BVH-11-2.  
XX OS Streptococcus pneumoniae.  
XX OS Synthetic.  
XX XX WO2003054007-A2.  
XX PN  
XX PD 03-JUL-2003.  
XX XX  
XX PF 20-DEC-2002; 2002WO-CA002006.  
XX XX 20-DEC-2001; 2001US-0341252P.  
XX PR  
XX XX (SHIR-) SHIRE BIOCHEM INC.  
XX PA  
XX PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;  
XX PI Labbe S;  
XX XX  
XX DR WPI; 2003-569224/53.  
XX XX  
XX PT New isolated polypeptides of Streptococcus pneumoniae, useful for  
XX PT diagnosing, preventing or treating streptococcal infection, meningitis,  
XX PT otitis media, bacteraemia or pneumonia infection.  
XX PS Example 1; SEQ ID NO 38; 79pp; English.  
XX CC The invention relates to a novel isolated polypeptide of Streptococcus  
XX CC pneumoniae. A polypeptide of the invention has antibacterial,  
XX CC antiinflammatory, and auditory activity, and is used as a vaccine. The  
XX CC polypeptide or pharmaceutical composition is useful for the prophylactic  
XX CC or therapeutic treatment of streptococcal infection, meningitis, otitis  
XX CC media, bacteraemia or pneumonia infection. The kit is useful for  
XX CC detecting or diagnosing streptococcal infection. The pharmaceutical  
XX CC composition is useful as a vaccine. The polynucleotides are useful in  
XX CC designing DNA probes for detecting circulating Streptococcus in a  
XX CC biological sample. The present sequence is used in the exemplification of  
XX CC the invention. Note: The sequence data for this patent is not fully  
XX CC represented in the printed specification, but is based on sequence  
XX CC information supplied by the European Patent Office  
XX SQ Sequence 690 AA;

Query Match 10.1%; Score 80; DB 7; Length 690;  
Best Local Similarity 100.0%; Pred. No. 8.8e-68;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 70  
DB 12 TVKENNRVSYIDGKQATKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90  
DB 72 GKVPYDAIISELLMKDPNY 91

RESULT 12  
AAB12766  
ID AAB12766 standard; protein; 821 AA.  
XX AC AAB12766;

CC described in the method of the invention. Note: This sequence does not  
CC appear in the specification but has been created according to information  
CC given in the invention

XX SQ Sequence 334 AA;

Query Match 10.1%; Score 80; DB 5; Length 334;  
Best Local Similarity 100.0%; Pred. No. 4.5e-68;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70  
Db 12 TVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90  
Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 8

ABM18808  
ID ABM18808 standard; protein; 334 AA.

XX AC ABM18808;

DT 13-OCT-2003 (first entry)

XX DE S. pneumoniae variant protein BVH-11A.

XX KW antibacterial; antiinflammatory; auditory; vaccine; meningitis;  
KW streptococcal infection; otitis media; bacteraemia; pneumonia; BVH-3;  
KW mutant; mutein; New 43; BVH-11; BVH-11-2.

XX OS Streptococcus pneumoniae.  
OS Synthetic.

PN WO2003054007-A2.

XX PD 03-JUL-2003.

XX PF 20-DEC-2002; 2002WO-CA002006.

XX PR 20-DEC-2001; 2001US-0341252P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Charland N, Brodeur BR, Martin D, Blais N, Ouellet C;  
PI Labbe S;

XX DR WPI; 2003-569224/53.

XX PT New isolated polypeptides of Streptococcus pneumoniae, useful for  
PT diagnosing, preventing or treating streptococcal infection, meningitis,  
PT otitis media, bacteraemia or pneumonia infection.

XX PS Example 1; SEQ ID NO 20; 79pp; English.

XX CC The invention relates to a novel isolated polypeptide of Streptococcus  
CC pneumoniae. A polypeptide of the invention has antibacterial,  
CC antiinflammatory, and auditory activity, and is used as a vaccine. The  
CC polypeptide or pharmaceutical composition is useful for the prophylactic  
CC or therapeutic treatment of streptococcal infection, meningitis, otitis  
CC media, bacteraemia or pneumonia infection. The kit is useful for  
CC detecting or diagnosing streptococcal infection. The pharmaceutical  
CC composition is useful as a vaccine. The polynucleotides are useful in  
CC designing DNA probes for detecting circulating Streptococcus in a  
CC biological sample. The present sequence is used in the exemplification of  
CC the invention. Note: The sequence data for this patent is not fully  
CC represented in the printed specification, but is based on sequence  
CC information supplied by the European Patent Office

XX SQ Sequence 334 AA;

Query Match 10.1%; Score 80; DB 7; Length 334;  
Best Local Similarity 100.0%; Pred. No. 4.5e-68;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70  
Db 12 TVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90  
Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 9

AAB12745  
ID AAB12745 standard; protein; 690 AA.

XX AC AAB12745;

XX DT 23-NOV-2000 (first entry)

XX DE Streptococcus pneumoniae NEW16 protein antigen SEQ ID NO:79.

XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal.

XX OS Streptococcus pneumoniae.

PN WO200039299-A2.

XX PD 06-JUL-2000.

XX PF 20-DEC-1999; 99WO-CA001218.

XX PR 23-DEC-1998; 98US-0113800P.

XX PA (BIOC-) BIOCHEM PHARMA INC.

XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX DR WPI; 2000-452397/39.

XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia.

XX PS Claim 18; Fig 44; 106pp; English.

XX CC The present invention describes nucleic acids (I) encoding protein  
XX antigens (II) from Streptococcus pneumoniae. The protein antigens have  
XX bactericidal activity. The nucleic acids, encoding the protein antigens,  
XX may be used for the recombinant production of the proteins they encode.  
XX The protein antigens may then be used as vaccines for the prevention and  
XX treatment of Streptococcal infections in mammals (especially humans)  
XX which result in, e.g. meningitis, otitis media, bacteraemia and/or  
XX pneumonia. The present sequence represents the S. pneumoniae NEW16  
XX protein antigen

XX SQ Sequence 690 AA;

Query Match 10.1%; Score 80; DB 3; Length 690;  
Best Local Similarity 100.0%; Pred. No. 8.8e-68;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70  
Db 12 TVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90  
Db 72 GKVPYDAIISELLMKDPNY 91

```
QY 421 VSHLTAKKNVAPRQDFYKAYNLLTEAHKALFXNKGNSDFQALDKLLRLNDESTN 480
DB 441 VSHLTAKKNVAPRQDFYKAYNLLTEAHKALFENKGRNSDFQALDKLLRLNDESTN 500
QY 481 KEKLVDDLLAFAPITHPERLGKPNQSOIYETEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
DB 501 KEKLVDDLLAFAPITHPERLGKPNQSOIYETEDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY 541 EGDAYVTHMGHSHWIGKDSLSDEKVAQAQYTKKGLPPSPADVKANPTGDSAAAIY 600
DB 561 EGDAYVTHMGHSHWIGKDSLSDEKVAQAQYTKKGLPPSPADVKANPTGDSAAAIY 620
QY 601 NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDVHNIKFAMFDDHTYKAPNGYTTLED 660
DB 621 NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDVHNIKFAMFDDHTYKAPNGYTTLED 680
QY 661 LFATIKYVVEHPDRPHNSDGNAGSEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720
DB 681 LFATIKYVVEHPDRPHNSDGNAGSEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 740
QY 721 QVTEKVEAQLKEAEVLLAKVTDSSLYKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 780
DB 741 QVTEKVEAQLKEAEVLLAKVTDSSLYKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLA 800
QY 781 LLKGSNPFSSVSKKIN 796
DB 801 LLKGSNPFSSVSKKIN 816

RESULT 6
AAB12728
ID AAB12728 standard; protein; 334 AA.
AC AAB12728;
XX
XX
XX 21-NOV-2000 (first entry)
DT
DE Streptococcus pneumoniae BVH-11A protein antigen SEQ ID NO:61.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
PN
PD
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA001218.
XX
XX 23-DEC-1998; 98US-0113800P.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
PA
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia.
XX
XX Claim 18; Fig 26; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens have
XX bactericidal activity. The nucleic acids, encoding the protein antigens,
XX may be used for the recombinant production of the proteins they encode.
XX The protein antigens may then be used as vaccines for the prevention and
XX treatment of Streptococcal infections in mammals (especially humans)
XX which result in, e.g. meningitis, otitis media, bacteraemia and/or
XX pneumonia. The present invention represents the S. pneumoniae BVH-11A
```

```
CC protein antigen
XX Sequence 334 AA;
SQ
Query Match 10.1%; Score 80; DB 3; Length 334;
Best Local Similarity 100.0%; Pred. No. 4.5e-68;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 TVKENRVSYDVGKQATKCTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGSHYHYN 70
DB 12 TVKENRVSYDVGKQATKCTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGSHYHYN 71
QY 71 GKVPYDAIISEELLMKDPNY 90
DB 72 GKVPYDAIISEELLMKDPNY 91

RESULT 7
AAU84027
ID AAU84027 standard; peptide; 334 AA.
XX
AC AAU84027;
XX
XX 08-MAY-2002 (first entry)
DT
XX Truncated variant of S. pneumoniae BVH-11, BVH-11A.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
XX streptococcal bacterial infection; mutant; mutein.
XX
XX Streptococcus pneumoniae.
OS
XX Synthetic.
XX
XX WO200198334-A2.
PN
XX 27-DEC-2001.
PD
XX
XX 19-JUN-2001; 2001WO-CA000908.
XX
XX 20-JUN-2000; 2000US-0212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
XX polypeptides, useful as vaccine components for treating or preventing
XX streptococcal infections such as otitis media, meningitis, and
XX bacteraemia.
XX
XX Example 1; Page; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90% identity
XX to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
XX BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
XX Staphylococcus aureus) in an individual susceptible to the infection. A
XX polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
```







XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;  
XX  
XX WPI; 2002-479261/51.  
DR N-PSDB; ABQ84819.  
XX  
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus  
PT and for preventing or attenuating disease caused by Streptococcus  
PT infection.  
XX  
XX Claim 11; Page 27; 70pp; English.  
XX  
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the  
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54569. The S.  
CC pneumoniae antigens have antibacterial activity and can be used in  
CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
CC attenuate a Streptococcus infection in an animal. The polynucleotides  
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning  
CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
CC from the present invention  
XX  
XX SQ Sequence 796 AA;  
  
Query Match 99.9%; Score 795; DB 5; Length 796;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60  
Db 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60  
  
QY 61 SHGDHYHYNGKVPYDAI1SEELLMDPNYKLDIEDIVNEVKGYVIVKDGKYYVYLKDA 120  
Db 61 SHGDHYHYNGKVPYDAI1SEELLMDPNYKLDIEDIVNEVKGYVIVKDGKYYVYLKDA 120  
  
QY 121 AHADNVRTKEENRQKQESQHQREGTTPRNDGVALARSQGRYTTDDGVIFNASDIIEDT 180  
Db 121 AHADNVRTKEENRQKQESQHQREGTTPRNDGVALARSQGRYTTDDGVIFNASDIIEDT 180  
  
QY 181 GDAYIVPHGDHYHYIPKNEISASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240  
Db 181 GDAYIVPHGDHYHYIPKNEISASELAFAFLSGRGNLSNRTYRQNSDNTSRTNWVPS 240  
  
QY 241 VSNPGTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDFDPAQITSTRT 300  
Db 241 VSNPGTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDFDPAQITSTRT 300  
  
QY 301 ARGVAVPHGDHYHYFIPYSQMSLEERARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360  
Db 301 ARGVAVPHGDHYHYFIPYSQMSLEERARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360  
  
QY 361 PQAPNLKIDSNSLSVSQLVRKVGEGVYFEEKISRYVFAKDLPSRTVKNLKSLKQES 420  
Db 361 PQAPNLKIDSNSLSVSQLVRKVGEGVYFEEKISRYVFAKDLPSRTVKNLKSLKQES 420  
  
QY 421 VSHLTAKENAVAPROEFYDKAYNLLTEAHKALFYKNGNSDFQALDKLLERLNDSTN 480  
Db 421 VSHLTAKENAVAPROEFYDKAYNLLTEAHKALFYKNGNSDFQALDKLLERLNDSTN 480  
  
QY 481 KEKLVDDLLAFLAPIITHPELGLKPNQSIETVEDEVIAQLADKYTTSDGYIFDEHDIISD 540  
Db 481 KEKLVDDLLAFLAPIITHPELGLKPNQSIETVEDEVIAQLADKYTTSDGYIFDEHDIISD 540  
  
QY 541 EGDAYVTPHMGSHWIGKOSLSKEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAIY 600  
Db 541 EGDAYVTPHMGSHWIGKOSLSKEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAIY 600  
  
QY 601 NRKVGKRIIPLVRLPYMVEHTVEKNGNLIIPKDHVHNKKEAFWDDHTYKAPNGYITLED 660  
Db 601 NRKVGKRIIPLVRLPYMVEHTVEKNGNLIIPKDHVHNKKEAFWDDHTYKAPNGYITLED 660

QY 661 LFATIKYVVEHDPDRPHSNDGWNASEHVLGKKHSDSDPNKFKADEEPVEETPAEPVP 720  
Db 661 LFATIKYVVEHDPDRPHSNDGWNASEHVLGKKHSDSDPNKFKADEEPVEETPAEPVP 720  
  
QY 721 QVETEKVRAQLKEAEVLLAKVTDSSIKANATETLAGLRNNLTLOIMDNNSINAAEAKLLA 780  
Db 721 QVETEKVRAQLKEAEVLLAKVTDSSIKANATETLAGLRNNLTLOIMDNNSINAAEAKLLA 780  
  
QY 781 LLKGSNPSSVSKEKIN 796  
Db 781 LLKGSNPSSVSKEKIN 796  
  
RESULT 3  
ADC45137  
ID ADC45137 standard; protein; 796 AA.  
XX  
AC ADC45137;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE S. pneumoniae antigenic protein SP036.  
XX  
XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US6573082-B1.  
XX  
PD 03-JUN-2003.  
XX  
PF 28-MAR-2000; 2000US-00536784.  
XX  
PR 31-OCT-1996; 96US-0029960P.  
PR 30-OCT-1997; 97US-00961083.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;  
XX  
XX WPI; 2003-764574/72.  
DR N-PSDB; ADC45136.  
XX  
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
PT useful for producing vaccines for prevention or attenuation of infection  
PT by Streptococcus pneumoniae.  
XX  
PS Example 1; SEQ ID NO 56; 58pp; English.  
XX  
XX The invention relates to an isolated polynucleotide consisting of a  
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
CC antigens. Also included are making a recombinant vector by inserting the  
CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
CC antigen nucleic acids are useful as probes for use in diagnostic methods  
CC for detecting S. pneumoniae gene expression. The present sequence  
CC represents an S. pneumoniae antigenic protein.  
XX  
SQ Sequence 796 AA;  
  
Query Match 99.9%; Score 795; DB 7; Length 796;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60  
Db 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60

```

XX AC AAW55090;
XX DT 02-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae SP0036 protein.
XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX OS detection; pneumonia; otitis media; meningitis.
XX ST Streptococcus pneumoniae.
XX FT Key Location/Qualifiers
XX FT Misc-difference 456
XX FT /label= unknown
XX FT /note= "encoded by GNA"
XX FN MO9818930-A2.
XX ST 07-MAY-1998.
XX PD 30-OCT-1997; 97WO-US019422.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR (HUMA-) HUMAN GENOME SCI INC.
XX PA Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX PI WPI; 1998-272224/24.
XX DR N-PSDB; AAV27351.
XX ST Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX PT - or their epitope-containing fragments, useful in protective or
XX PT therapeutic vaccines, and for diagnosis.
XX ST Claim 11; Page 59-60; 118pp; English.
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX CC can be useful in vaccines for inducing protective antibodies against
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX CC are used to detect Streptococcus infection (by usual hybridisation or
XX CC amplification methods), also for isolating Streptococcus genes or their
XX CC allelic variants. The protein can be used similarly to detect specific
XX CC antibodies in standard immunoassays, especially for diagnosing or
XX CC monitoring infections. Antibodies which bind the protein are used to
XX CC detect corresponding antigens, to purify the protein and for passive
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX CC (especially 10-300) mu g/ml per dose
XX ST Sequence 796 AA;
XX SQ
Query Match 99.9%; Score 795; DB 2; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYELGLYQARTVKNRVSVIDGKATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 60
DB 1 SYELGLYQARTVKNRVSVIDGKATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKYPYDAIISSEILLMDPNYKLDIEDIVNEVKGGYIVKDGKYYVYLKDA 120
DB 61 SHGDHYHYNGKYPYDAIISSEILLMDPNYKLDIEDIVNEVKGGYIVKDGKYYVYLKDA 120
QY 121 AHADNVRTKEEINRQKQESHQREGTTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180
DB 121 AHADNVRTKEEINRQKQESHQREGTTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGNLSRSTYRRQNSDNTSRTNWVPS 240

```

```

DB 181 GDAYIVPHGDHYHYIPKNELSASELAFAFLSGRGNLSRSTYRRQNSDNTSRTNWVPS 240
QY 241 VSNPQTNTNTSNNSTNSQASQNDIDSLKQLYKPLSORHVESDGLVDFDPAQITSR 300
DB 241 VSNPQTNTNTSNNSTNSQASQNDIDSLKQLYKPLSORHVESDGLVDFDPAQITSR 300
QY 301 ARGVAVPHGDHYHYIPYSQMSSELEERIARIIPLYRSNHNWVPDSRPEQSPQPTPEPSG 360
DB 301 ARGVAVPHGDHYHYIPYSQMSSELEERIARIIPLYRSNHNWVPDSRPEQSPQPTPEPSG 360
QY 361 POPAPNLKIDNSLSLVOLVKVGEVYVREKIGISRVYFAKDLPSFTVKLESKLSQES 420
DB 361 POPAPNLKIDNSLSLVOLVKVGEVYVREKIGISRVYFAKDLPSFTVKLESKLSQES 420
QY 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
DB 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLLAFLAPITHPERLKGPNISQIETEDVRITAOADKYTTSDGYIFDEHDIISD 540
DB 481 KEKLVDDLLAFLAPITHPERLKGPNISQIETEDVRITAOADKYTTSDGYIFDEHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKGILPPSPDADVKNPTGDSAAIY 600
DB 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYTKGILPPSPDADVKNPTGDSAAIY 600
QY 601 NRVKGEKRIPLVRLPYMVEHTVEVQGNLIIPHKHVHNKIFAFDDHTYKAPNGYTTLED 660
DB 601 NRVKGEKRIPLVRLPYMVEHTVEVQGNLIIPHKHVHNKIFAFDDHTYKAPNGYTTLED 660
QY 661 LFATIKYVVEHPDHPHSDNGWGNASEHVLGKHSDPNKPKADEEPVEETPAEPEVP 720
DB 661 LFATIKYVVEHPDHPHSDNGWGNASEHVLGKHSDPNKPKADEEPVEETPAEPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLKANATETLAGLRNNLTQIMDNNSIMAEKLLA 780
DB 721 QVETEKVEAQLKEAEVLLAKVTDSLKANATETLAGLRNNLTQIMDNNSIMAEKLLA 780
QY 781 LLKGSNPSSVSKEKIN 796
DB 781 LLKGSNPSSVSKEKIN 796
RESULT 2
ABP54584
ID ABP54584 standard; protein; 796 AA.
XX AC ABP54584;
XX XX
XX DT 04-SEP-2002 (first entry)
XX DE S. pneumoniae SP036 protein sequence SEQ ID NO:56.
XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX KW antibacterial; Streptococcal infection; detection.
XX OS Streptococcus pneumoniae.
XX PN US2002061545-A1.
XX PD 23-MAY-2002.
XX PF 22-JAN-2001; 2001US-00765272.
XX PR 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:17:24 ; Search time 81 Seconds  
(without alignments)  
2776.640 Million cell updates/sec

Title: US-09-765-271-56  
Perfect score: 796  
Sequence: 1 SYELGLQARTVKNRVS.....KLLALLKGSNPFSSVSKKIN 796

Scoring table: OMIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : A.Geneseq.29Jan04.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1980s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	2	AAW55090 Streptococ
2	795	99.9	796	5	ABP54584 S. pneumo
3	795	99.9	796	5	ADC45137 S. pneumo
4	695	87.3	802	6	ABU01598 S. pneumo
5	695	87.3	819	3	AAU01468 Recombina
6	80	10.1	334	3	AAU01468 Recombina
7	80	10.1	334	5	AAU04027 Truncated
8	80	10.1	334	7	ABM18808 S. pneumo
9	80	10.1	690	3	AAU012745 Streptoco
10	80	10.1	690	5	AAU04093 Truncated
11	80	10.1	690	7	ABM18826 S. pneumo
12	80	10.1	821	3	AAU012766 Streptoco
13	80	10.1	821	3	AAU012727 Streptoco
14	80	10.1	821	5	AAU04026 Truncated
15	80	10.1	821	7	ABM18807 S. pneumo
16	80	10.1	840	3	AAU012716 Streptoco
17	80	10.1	840	5	AAU075933 Streptoco
18	80	10.1	840	7	ABM18797 S. pneumo
19	60	7.5	805	3	AAU012764 Streptoco
20	60	7.5	807	3	AAU012765 Streptoco
21	60	7.5	811	3	AAU012761 Streptoco
22	60	7.5	811	3	AAU012762 Streptoco
23	60	7.5	811	3	AAU012760 Streptoco
24	60	7.5	811	3	AAU012763 Streptoco
25	60	7.5	816	3	AAU012758 Streptoco

26	60	7.5	816	3	AAU012757 Streptoco
27	60	7.5	816	3	AAU012756 Streptoco
28	60	7.5	819	3	AAU012754 Streptoco
29	60	7.5	819	3	AAU012740 Streptoco
30	60	7.5	819	3	AAU01469 Recombina
31	60	7.5	819	5	AAU04087 Truncated
32	60	7.5	819	6	ABU01597 S. pneumo
33	60	7.5	819	7	ABM18820 S. pneumo
34	60	7.5	820	3	AAU012755 Streptoco
35	60	7.5	820	3	AAU012755 Streptoco
36	60	7.5	827	3	AAU012759 Streptoco
37	60	7.5	834	3	AAU012759 Streptoco
38	60	7.5	838	3	AAU012720 Streptoco
39	60	7.5	838	3	AAU01466 Recombina
40	60	7.5	838	5	AAU075934 Streptoco
41	60	7.5	838	7	ABM18798 S. pneumo
42	60	7.5	839	6	ABU01418 S. pneumo
43	56	7.0	613	3	AAU012730 Streptoco
44	56	7.0	613	5	AAU04029 Truncated
45	56	7.0	613	7	ABM18810 S. pneumo
46	51	6.4	721	2	AAU05753 Streptoco
47	51	6.4	763	2	AAU055095 Streptoco
48	51	6.4	763	5	ABP54589 S. pneumo
49	51	6.4	763	7	ADC45147 S. pneumo
50	32	4.0	42	5	AAU083828 S. pneumo
51	32	4.0	205	3	AAU012726 Streptoco
52	32	4.0	205	5	AAU04025 Truncated
53	32	4.0	205	7	ABM18806 S. pneumo
54	32	4.0	447	2	AAU061228 Streptoco
55	32	4.0	447	5	ABP54647 S. pneumo
56	32	4.0	447	7	ADC45263 S. pneumo
57	32	4.0	484	3	AAU012718 Streptoco
58	32	4.0	484	3	AAU01467 Recombina
59	32	4.0	485	3	AAU01538 Streptoco
60	32	4.0	489	3	AAU012723 Streptoco
61	32	4.0	489	5	AAU04022 Truncated
62	32	4.0	489	7	ABM18803 S. pneumo
63	32	4.0	489	7	ABM18803 S. pneumo
64	32	4.0	509	3	AAU012724 Streptoco
65	32	4.0	509	5	AAU04023 Truncated
66	32	4.0	509	7	ABM18804 S. pneumo
67	32	4.0	679	5	AAU04092 Truncated
68	32	4.0	780	3	AAU012744 Streptoco
69	32	4.0	780	7	ABM18825 S. pneumo
70	32	4.0	840	3	AAU012721 Streptoco
71	32	4.0	840	5	AAU076151 Streptoco
72	32	4.0	1019	3	AAU012749 Streptoco
73	32	4.0	1019	3	AAU012753 Streptoco
74	32	4.0	1019	3	AAU012722 Streptoco
75	32	4.0	1019	3	AAU012751 Streptoco
76	32	4.0	1019	3	AAU012750 Streptoco
77	32	4.0	1019	3	AAU012748 Streptoco
78	32	4.0	1019	5	AAU04021 Truncated
79	32	4.0	1019	7	ABM18802 S. pneumo
80	32	4.0	1039	3	AAU012715 Streptoco
81	32	4.0	1039	5	AAU075932 Streptoco
82	32	4.0	1039	6	ABU01419 S. pneumo
83	32	4.0	1039	7	ABM18796 S. pneumo
84	32	4.0	1039	7	ABM18827 S. pneumo
85	27	3.4	294	5	AAU04094 Truncated
86	27	3.4	294	7	ABM18827 S. pneumo
87	27	3.4	466	5	ABP26208 Streptoco
88	27	3.4	473	3	AAU012743 Streptoco
89	27	3.4	473	5	AAU04091 Truncated
90	27	3.4	473	7	ABM18824 S. pneumo

## ALIGNMENTS

RESULT 1  
AAW55090 standard; protein: 796 AA.  
ID AAW55090

```

US-10-387-783-1
; Sequence 1, Application US/10387783
; Publication No. US2004000531A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Forward primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-10-387-783-1

```

```

Query Match      1.1%; Score 27; DB 16; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
    |||||
Db 10 TTCTTACGAGTTGGGACTGTATCAAGC 36

```

```

Search completed: October 1, 2004, 07:13:14
Job time : 1914 secs

```





; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-412-850-5

Query Match 2.2%; Score 53; DB 16; Length 2531;  
Best Local Similarity 100.0%; Pred. No. 1e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1035  
|||||  
DB 1003 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1055

## RESULT 14

US-10-412-850-11  
; Sequence 11, Application US/10412850  
; Publication No. US20040001836A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie E.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-686  
; CURRENT APPLICATION NUMBER: US/10/412,850  
; CURRENT FILING DATE: 2003-04-14  
; PRIOR APPLICATION NUMBER: 09/468,656  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/113,048  
; PRIOR FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 11  
; LENGTH: 2531  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-412-850-11

Query Match 2.2%; Score 53; DB 16; Length 2531;  
Best Local Similarity 100.0%; Pred. No. 1e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1035  
|||||  
DB 1000 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1052

## RESULT 15

US-10-387-783-5  
; Sequence 5, Application US/10387783  
; Publication No. US20040005331A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie E.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-683  
; CURRENT APPLICATION NUMBER: US/10/387,783  
; CURRENT FILING DATE: 2003-03-13  
; PRIOR APPLICATION NUMBER: 09/468,656  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/113,048  
; PRIOR FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 5  
; LENGTH: 2531  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-387-783-5

Query Match 2.2%; Score 53; DB 16; Length 2531;

Best Local Similarity 100.0%; Pred. No. 1e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1035  
|||||  
DB 1003 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1055

## RESULT 16

US-10-387-783-11  
; Sequence 11, Application US/10387783  
; Publication No. US20040005331A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie E.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-683  
; CURRENT APPLICATION NUMBER: US/10/387,783  
; CURRENT FILING DATE: 2003-03-13  
; PRIOR APPLICATION NUMBER: 09/468,656  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/113,048  
; PRIOR FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 11  
; LENGTH: 2531  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-387-783-11

Query Match 2.2%; Score 53; DB 16; Length 2531;  
Best Local Similarity 100.0%; Pred. No. 1e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1035  
|||||  
DB 1000 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1052

## RESULT 17

US-09-884-465A-5  
; Sequence 5, Application US/09884465A  
; Publication No. US20030077293A1  
; GENERAL INFORMATION:  
; APPLICANT: Shire Biochem, Inc.  
; APPLICANT: Hamel, Josee  
; APPLICANT: Brodeur, Bernard  
; APPLICANT: Martin, Denis  
; APPLICANT: Charland, Nathalie  
; APPLICANT: Ouellet, Catherine  
; TITLE OF INVENTION: Streptococcus Antigens  
; FILE REFERENCE: 055190-0044  
; CURRENT APPLICATION NUMBER: US/09/884,465A  
; CURRENT FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 60/212,683  
; PRIOR FILING DATE: 2000-06-20  
; NUMBER OF SEQ ID NOS: 384  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 2639  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-09-884-465A-5

Query Match 2.2%; Score 53; DB 10; Length 2639;  
Best Local Similarity 100.0%; Pred. No. 1e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGTACCAGATTCAAG 1035  
|||||



```
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hyman, Mark J.
;   REGISTRATION NUMBER: 46,789
;   REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 243:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2359 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 243:
US-10-158-844-243

Query Match      2.2%; Score 53; DB 13; Length 2359;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1879 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1931

RESULT 10
US-09-769-787-206
; Sequence 206, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-206

Query Match      2.2%; Score 53; DB 10; Length 2481;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1055

RESULT 11
US-10-412-862-5
; Sequence 5, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
```

```
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-5

Query Match      2.2%; Score 53; DB 13; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1055

RESULT 12
US-10-412-862-11
; Sequence 11, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-11

Query Match      2.2%; Score 53; DB 13; Length 2531;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035
Db 1000 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1052

RESULT 13
US-10-412-850-5
; Sequence 5, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
```

NUMBER OF SEQ ID NOS: 384  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 2523  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-09-884-465A-3

Query Match 6.2%; Score 148; DB 10; Length 2523;  
Best Local Similarity 100.0%; Pred. No. 3.3e-67;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 699 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTACAATATCTAACACAAGCA 758  
Db 758 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTACAATATCTAACACAAGCA 817  
QY 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAGTAAATGACATTGATGTCTCTTGAAC 818  
Db 818 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAGTAAATGACATTGATGTCTCTTGAAC 877  
QY 819 AGCTCTACAACTGCCTTTGAGTCAACG 846  
Db 878 AGCTCTACAACTGCCTTTGAGTCAACG 905

## RESULT 7

US-09-884-465A-4  
Sequence 4, Application US/09884465A  
Publication No. US20030077293A1  
GENERAL INFORMATION:  
APPLICANT: Shire Biochem, Inc.  
APPLICANT: Hamel, Josee  
APPLICANT: Brodeur, Bernard  
APPLICANT: Martin, Denis  
APPLICANT: Charland, Nathalie  
APPLICANT: Ouellet, Catherine  
TITLE OF INVENTION: Streptococcus Antigens  
FILE REFERENCE: 055190-0044  
CURRENT APPLICATION NUMBER: US/09/884,465A  
CURRENT FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: 60/212,683  
PRIOR FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 384  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 2647  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-09-884-465A-4

Query Match 6.2%; Score 148; DB 10; Length 2647;  
Best Local Similarity 100.0%; Pred. No. 3.4e-67;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 699 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTACAATATCTAACACAAGCA 758  
Db 802 CAAGAACAACTGGGTACCTCTGTAAGCAATCCAGGAAGTACAATATCTAACACAAGCA 861  
QY 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAGTAAATGACATTGATGTCTCTTGAAC 818  
Db 862 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAGTAAATGACATTGATGTCTCTTGAAC 921  
QY 819 AGCTCTACAACTGCCTTTGAGTCAACG 846  
Db 922 AGCTCTACAACTGCCTTTGAGTCAACG 949

## RESULT 8

US-09-765-272-65  
Sequence 65, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2290 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 65:  
US-09-765-272-65

Query Match 2.2%; Score 53; DB 9; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 1e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAAG 1035  
Db 944 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAAG 996

## RESULT 9

US-10-158-844-243  
Sequence 243, Application US/10158844  
Publication No. US20040029118A1  
GENERAL INFORMATION:

APPLICANT: Kunsch et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
CORRESPONDENCE ADDRESS:  
NUMBER OF SEQUENCES: 391

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude Pentium 3  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/158,844  
FILING DATE: 03-Jun-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/961,527  
FILING DATE: 1997-10-30

QY 481 AGGACGCTATATCAAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGGATAC 540  
Db 3533 AGGACGCTATATCAAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGGATAC 3532  
QY 541 TGGTGATGCTTATATCGTTTCTCATGGAGATCAATACCATTTACATTTCTTAAAGATGAGTT 600  
Db 3593 TGGTGATGCTTATATCGTTTCTCATGGAGATCAATACCATTTACATTTCTTAAAGATGAGTT 3652  
QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAGAGCTTCTATCTGCTGAGAGAAATCTGTCAA 660  
Db 3653 ATCAGCTAGCGAGTTGGCTGCTGAGAGAGCTTCTATCTGCTGAGAGAAATCTGTCAA 3712  
QY 661 TTCAGAAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720  
Db 3713 TTCAGAAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 3772  
QY 721 TGTAAAGCAATCCAGAACTACAAATATCTAACAAGCAAAACAAAGCAAACTAACAAGTCA 780  
Db 3773 TGTAAAGCAATCCAGAACTACAAATATCTAACAAGCAAAACAAAGCAAACTAACAAGTCA 3832  
QY 781 AGCAAGTCMAAGTAAATGACATTTGATGCTCTTGAACAGCTCTACAACTGCCCTTTGAG 840  
Db 3833 AGCAAGTCMAAGTAAATGACATTTGATGCTCTTGAACAGCTCTACAACTGCCCTTTGAG 3892  
QY 841 TCAACGACATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 900  
Db 3893 TCAACGACATGTAGAATCTGATGCCCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 3952  
QY 901 AGCTAGAGGTGTCAGTGCACAGGAGATCATTAACCTTATCCCTTACTCTCAAAAT 960  
Db 3953 AGCTAGAGGTGTCAGTGCACAGGAGATCATTAACCTTATCCCTTACTCTCAAAAT 4012  
QY 961 GTCTGAATTTGGAAGAAAGCAATGCTCTGATTTATTTCCCTTCTGTTATCTGTTCAAAACCATTTG 1020  
Db 4013 GTCTGAATTTGGAAGAAAGCAATGCTCTGATTTATTTCCCTTCTGTTATCTGTTCAAAACCATTTG 4072  
QY 1021 GGTAACGATTTAAGGCGAGAAACCAAGTCCAACCGACTCCGGAACCTAGTCCAGG 1080  
Db 4073 GGTAACGATTTAAGGCGAGAAACCAAGTCCAACCGACTCCGGAACCTAGTCCAGG 4132  
QY 1081 CCCGCACTGCGACCAATCTTAAATAGACTCAAAATCTTCTTGTGTAGTCAAGTGGT 1140  
Db 4133 CCCGCACTGCGACCAATCTTAAATAGACTCAAAATCTTCTTGTGTAGTCAAGTGGT 4192  
QY 1141 ACAGAAAGTTGGGAGAGATATGATTTGGAAGAAAGGGCATCTCTGTTATGCTTTTGC 1200  
Db 4193 ACAGAAAGTTGGGAGAGATATGATTTGGAAGAAAGGGCATCTCTGTTATGCTTTTGC 4252  
QY 1201 GAAAGATTTACCAATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1260  
Db 4253 GAAAGATTTACCAATCTGAAACTGTTAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 4312  
QY 1261 TGTTCACACATTTAACTCTAAGAAAGAAATGTTGCTCCTCGTGACCAAGATTTTA 1320  
Db 4313 TGTTCACACATTTAACTCTAAGAAAGAAATGTTGCTCCTCGTGACCAAGATTTTA 4372  
QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTCGNAATTAAGGCTG 1380  
Db 4373 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTCGNAATTAAGGCTG 4432  
QY 1381 TAATCTGATTTCCAGCCTTAGACAAATATTATAGAACCGTTGAATGATGAATCGACTAA 1440  
Db 4433 TAATCTGATTTCCAGCCTTAGACAAATATTATAGAACCGTTGAATGATGAATCGACTAA 4492  
QY 1441 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCAACCAATACCCATCCAGCG 1500  
Db 4493 TAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCAACCAATACCCATCCAGCG 4552  
QY 1501 ACTTGGCAAAACCAAAATCTCAAAATGAGTATATCTGAAGCAAGATTTGCTTCTCAAT 1560  
Db 4553 ACTTGGCAAAACCAAAATCTCAAAATGAGTATATCTGAAGCAAGATTTGCTTCTCAAT 4612  
QY 1561 AGCTGATAAGTATACAAAGTCAGATGGTTACATTTTGTGTAACATGATATATCATGTA 1620

Db 4613 AGCTGATAAGTATACAAAGTCAGATGGTTACATTTTGTGTAACATGATATATCATGTA 4672  
QY 1621 TGAAGGAGATGCATATGATAAGCGCTCATATGGGCCATAGTACACTGGATTTGAAAAAGTAG 1680  
Db 4673 TGAAGGAGATGCATATGATAAGCGCTCATATGGGCCATAGTACACTGGATTTGAAAAAGTAG 4732  
QY 1681 CCTTCTGATAAGAAAAAGTTGAGCTCAAGCTTATCTAAGAAAAAGTATCTTACC 1740  
Db 4733 CCTTCTGATAAGAAAAAGTTGAGCTCAAGCTTATCTAAGAAAAAGTATCTTACC 4792  
QY 1741 TCCATCTCCAGACGACATGTTTAAAGCAAAATCCAACCTGGAGATAGTCCAGCAGCTTTTA 1800  
Db 4793 TCCATCTCCAGACGACATGTTTAAAGCAAAATCCAACCTGGAGATAGTCCAGCAGCTTTTA 4852  
QY 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860  
Db 4853 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 4912  
QY 1861 TACAGTTGAGGTTTAAAAAACGTTAATTTGATTTATCTCTATAAGGATCATTTACCAATAT 1920  
Db 4913 TACAGTTGAGGTTTAAAAAACGTTAATTTGATTTATTTCTCTATAAGGATCATTTACCAATAT 4972  
QY 1921 TAAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980  
Db 4973 TAAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGA 5032  
QY 1981 TTTGTTTGGCAGATTAAGTACTACGTAGAACCCCTGACGAAACGCTCCACATTTCTATGA 2040  
Db 5033 TTTGTTTGGCAGATTAAGTACTACGTAGAACCCCTGACGAAACGCTCCACATTTCTATGA 5092  
QY 2041 TGGATGGGCAATGCCAGTCAGCATGTTTAGCAAGAAACACCTGCTGAGCCAGAGTCCC 2100  
Db 5093 TGGATGGGCAATGCCAGTCAGCATGTTTAGCAAGAAACACCTGCTGAGCCAGAGTCCC 5152  
QY 2101 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
Db 5153 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 5212  
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGAGAGTTTGTTCGAA 2220  
Db 5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGAGAGTTTGTTCGAA 5272  
QY 2221 AGTAACGAGTTCTAGTCTGAAGCCCAATGCCAGAAACCTCTAGCTGGTTTACGAATAA 2280  
Db 5273 AGTAACGAGTTCTAGTCTGAAGCCCAATGCCAGAAACCTCTAGCTGGTTTACGAATAA 5332  
QY 2281 TTTGACTCTTCAAAATATGATAAACAATAGTATCATGGCAGAGAGAGAGTTTACTTGC 2340  
Db 5333 TTTGACTCTTCAAAATATGATAAACAATAGTATCATGGCAGAGAGAGAGTTTACTTGC 5392  
QY 2341 GTTGTAAAGGAGTAATCTTCTATCTGTAAAGGAAAAATAAAC 2389  
Db 5393 GTTGTAAAGGAGTAATCTTCTATCTGTAAAGGAAAAATAAAC 5441

## RESULT 6

US-09-884-465A-3  
; Sequence 3, Application US/09884465A  
; Publication No. US20030077293A1  
; GENERAL INFORMATION:  
; APPLICANT: Shire Biochem, Inc.  
; APPLICANT: Hamel, Josee  
; APPLICANT: Brodeur, Bernard  
; APPLICANT: Martin, Denis  
; APPLICANT: Charland, Nathalie  
; APPLICANT: Ouellet, Catherine  
; TITLE OF INVENTION: Streptococcus Antigens  
; FILE REFERENCE: 055190-0044  
; CURRENT APPLICATION NUMBER: US/09/884,465A  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 60/212,683  
; PRIOR FILING DATE: 2000-06-20

Db	1440	TAATTTCTGATTTTCCAAAGCCTTAGACAAATTTATTAGAACCTTGAATGATGAATCGACTAA	1499
Qy	1441	TAAAGAAAAATTTGGTAGATGATTTATTGGCATTTCTTAGCACCAATTAATPACCATCCAGAGCG	1500
Db	1500	TAAAGAAAAATTTGGTAGATGATTTATTGGCATTTCTTAGCACCAATTAATPACCATCCAGAGCG	1559
Qy	1501	ACTTGGCAAAACCAAAATTTCTAAATTTGAGTATATCTGAAGACGAAGTTTCGTATTGCTCAATTT	1560
Db	1560	ACTTGGCAAAACCAAAATTTCTAAATTTGAGTATATCTGAAGACGAAGTTTCGTATTGCTCAATTT	1619
Qy	1561	AGCTGATAAGTATACAAACGTCAGATGGTTACATTTTGGATGAACATGATATAATCAGTGA	1620
Db	1620	AGCTGATAAGTATACAAACGTCAGATGGTTACATTTTGGATGAACATGATATAATCAGTGA	1679
Qy	1621	TGAAGGAGATGTCATATGTATTAACGGCTCATATGGGCCATAGTCACTGGATTTGGAAAAAGATAG	1680
Db	1680	TGAAGGAGATGTCATATGTATTAACGGCTCATATGGGCCATAGTCACTGGATTTGGAAAAAGATAG	1739
Qy	1681	CTTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACCTAAGAAAAAGGTATCTCTACC	1740
Db	1740	CTTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACCTAAGAAAAAGGTATCTCTACC	1799
Qy	1741	TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA	1800
Db	1800	TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA	1859
Qy	1801	CAATTCGTGTGAAGGGGAAAAAGAAATTCACCTCGTTTCGACTTCCTCATATATGTTTCAGCA	1860
Db	1860	CAATTCGTGTGAAGGGGAAAAAGAAATTCACCTCGTTTCGACTTCCTCATATATGTTTCAGCA	1919
Qy	1861	TACAGTTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAAAGGATCAATTAACCATTAATAT	1920
Db	1920	TACAGTTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAAAGGATCAATTAACCATTAATAT	1979
Qy	1921	TAAATTTTCTGTTGTTGATGATCACATACAAAGTTCCTCAATTCCTCATTAACCTTGGGAAGA	1980
Db	1980	TAAATTTTCTGTTGTTGATGATCACATACAAAGTTCCTCAATTCCTCATTAACCTTGGGAAGA	2039
Qy	1981	TTTGTGTTGCGACGATTAAGTACTACGTAGAACACCTTGACGAAGCTCCACATTTCTAATGA	2040
Db	2040	TTTGTGTTGCGACGATTAAGTACTACGTAGAACACCTTGACGAAGCTCCACATTTCTAATGA	2099
Qy	2041	TGGATGGGGCAATGCCAGTGAAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA	2100
Db	2100	TGGATGGGGCAATGCCAGTGAAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA	2159
Qy	2101	TAAAGATTTCAAACGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	2160	TAAAGATTTCAAACGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2219
Qy	2161	TCAAGTATAGACTCAAAAAGTTAGAAAGCCCAACTCAAAGACGAGAAGTTTGTGTTGCGAA	2220
Db	2220	TCAAGTATAGACTCAAAAAGTTAGAAAGCCCAACTCAAAGACGAGAAGTTTGTGTTGCGAA	2279
Qy	2221	AGTAAACGGATTTCTAGTCTGAAGCCAAATGCAACAGAAACCTCTAGCTGGTTTACGAATAAA	2280
Db	2280	AGTAAACGGATTTCTAGTCTGAAGCCAAATGCAACAGAAACCTCTAGCTGGTTTACGAATAAA	2339
Qy	2281	TTTTCAGCTCTTCAAATTTATGGATACAAATAGTATCATGTGCAGAACGAAAAATTTACTTGC	2340
Db	2340	TTTTCAGCTCTTCAAATTTATGGATACAAATAGTATCATGTGCAGAACGAAAAATTTACTTGC	2399
Qy	2341	GTGTGTTAAAGGAAAGTAAATCTCTCATCTGTAAGTAAGAAAAATAAAC	2389
Db	2400	GTGTGTTAAAGGAAAGTAAATCTCTCATCTGTAAGTAAGAAAAATAAAC	2448

APPLICANT: Kunsch et al.  
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 NUMBER OF SEQUENCES: 391  
 CORRESPONDENCE ADDRESS:  
     ADDRESSEE: Human Genome Sciences, Inc.  
     STREET: 9410 Key West Avenue  
     CITY: Rockville  
     STATE: Maryland  
     COUNTRY: USA  
     ZIP: 20850  
 COMPUTER READABLE FORM:  
     MEDIUM TYPE: CD-R  
     COMPUTER: Dell Latitude Pentium 3  
     OPERATING SYSTEM: Windows 98  
     SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
     APPLICATION NUMBER: US/10/158,844  
     FILING DATE: 03-Jun-2002  
     CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
     APPLICATION NUMBER: US 08/961,527  
     FILING DATE: 1997-10-30  
     APPLICATION NUMBER: US 60/029,960  
     FILING DATE: 1996-10-31  
 ATTORNEY/AGENT INFORMATION:  
     NAME: Hyman, Mark J.  
     REGISTRATION NUMBER: 46,789  
     REFERENCE/DOCKET NUMBER: PB340PID1  
 INFORMATION FOR SEQ ID NO: 94:  
     SEQUENCE CHARACTERISTICS:  
         LENGTH: 8195 base pairs  
         TYPE: nucleic acid  
         STRANDEDNESS: double  
         TOPOLOGY: linear  
     SEQUENCE DESCRIPTION: SEQ ID NO: 94:

Query Match	97.9%; Score 2338; DB 13; Length 8195;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2388; Conservative 0;	Mismatches 1; Indels 0; Gaps 0;
Qy	1 TTCTTAGCGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAAATAATCGTGTTCCTA 60
Db	3053 TTCTTAGCGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAAATAATCGTGTTCCTA 3112
Qy	61 TATAGATGGAAAAACAAGCGACGCAAAAACGGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
Db	3113 TATAGATGGAAAAACAAGCGACGCAAAAACGGAGAAATTTGACTCCTGATGAGGTTAGCAA 3172
Qy	121 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAACCAAGGCTATGTCC 180
Db	3173 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAACCAAGGCTATGTCC 3232
Qy	181 TTCACATGGCGCACCATATCATTTACAAATGGTAAAGTTCCTTATGACGCTATCATCAG 240
Db	3233 TTCACATGGCGCACCATATCATTTACAAATGGTAAAGTTCCTTATGACGCTATCATCAG 3292
Qy	241 TGAAGAAATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGGATATGTTAATGA 300
Db	3293 TGAAGAAATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGGATATGTTAATGA 3352
Qy	301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTAAGGATGC 360
Db	3353 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAATCTATGTTTACCTTAAGGATGC 3412
Qy	361 TGCCACGCGGATACGTCCGTACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
Db	3413 TGCCACGCGGATACGTCCGTACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 3472
Qy	421 TCAACATCGTGAAGGTGGAATCTCCAGAAAAACGATGGTGTGTCCTTGGCAGCTTCGCA 480
Db	3473 TCAACATCGTGAAGGTGGAATCTCCAGAAAAACGATGGTGTGTCCTTGGCAGCTTCGCA 3532

RESULT 5  
US-10-158-844-94  
; Sequence 94, Application US/10158844  
; Publication NO. US20040029118A1  
; GENERAL INFORMATION:

```

Db 2160 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCCAGAACTCCC 2219
QY 2161 TCAAGTAGAGACTGAAGAACTAGAGCCCAACTCAAGAGACGAGAACTTTTGGTTGGCAA 2220
Db 2220 TCAAGTAGAGACTGAAGAACTAGAGCCCAACTCAAGAGACGAGAACTTTTGGTTGGCAA 2279
QY 2221 AGTAACGGATTCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
Db 2280 AGTAACGGATTCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2339
QY 2281 TTTGACTCTTCAAAATTATGGAATAACAATAGTATCATGCGAGAGACGAGAAATAATTACTTGC 2340
Db 2340 TTTGACTCTTCAAAATTATGGAATAACAATAGTATCATGCGAGAGACGAGAAATAATTACTTGC 2399
QY 2341 GTTCTTAAAGGAGTAACTCTTCACTCTGTAAGTAAGGAAATAATAAC 2389
Db 2400 GTTCTTAAAGGAGTAACTCTTCACTCTGTAAGTAAGGAAATAATAAC 2448

RESULT 4
US-10-387-783-9
; Sequence 9, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; PRIOR FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-387-783-9

Query Match 100.0%; Score 2389; DB 16; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTCTTACGAGTGGGACTGTATCAAGCTAGACGAGTAAAGGAAATAATCGTTTTCCTA 60
Db 60 TTTCTTACGAGTGGGACTGTATCAAGCTAGACGAGTAAAGGAAATAATCGTTTTCCTA 119
QY 61 TATAGATGGAAGAACGAGGACGCAAAACCGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
Db 120 TATAGATGGAAGAACGAGGACGCAAAACCGAGAAATTTGACTCCTGATGAGGTTAGCAA 179
QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGCAAGGCTATGTCAC 180
Db 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGCAAGGCTATGTCAC 239
QY 181 TTCACATGGCGACCACTATCATTTATTAATGATGTAAGGTTCCCTTATGAGCTATCATCAG 240
Db 240 TTCACATGGCGACCACTATCATTTATTAATGATGTAAGGTTCCCTTATGAGCTATCATCAG 299
QY 241 TGAAGAAATTACTCATGAAGATCAAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300
Db 300 TGAAGAAATTACTCATGAAGATCAAACTATAAGCTAAAGATGAGGATATTGTTAATGA 359

```

```

QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db 360 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 419
QY 361 TGCCCAACGGGATAAAGTCCGTACAAAGAGGAAATCAATCGACAAATAAAGAGCATAG 420
Db 420 TGCCCAACGGGATAAAGTCCGTACAAAGAGGAAATCAATCGACAAATAAAGAGCATAG 479
QY 421 TCAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGTGTTGCCCTTGGCAGCTTCGCA 480
Db 480 TCAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGTGTTGCCCTTGGCAGCTTCGCA 539
QY 481 AGGACGCTTACTACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGGATAC 540
Db 540 AGGACGCTTACTACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGGATAC 599
QY 541 TGTGTGATGCTTATATCGTTCCTCATGAGATCATTAACCATTAATTCCTTAAGATGATTT 600
Db 600 TGTGTGATGCTTATATCGTTCCTCATGAGATCATTAACCATTAATTCCTTAAGATGATTT 659
QY 601 ATCAGCTAGCGAGTGGCTGCTGAGAGGCTTCTTATCTGGTCGAGAAATCTGTCAAA 660
Db 660 ATCAGCTAGCGAGTGGCTGCTGAGAGGCTTCTTATCTGGTCGAGAAATCTGTCAAA 719
QY 661 TTCAAGAACTTATCGCGACAAATAAGCGATACACTTCAAGAACAACTGGGTACCTTC 720
Db 720 TTCAAGAACTTATCGCGACAAATAAGCGATACACTTCAAGAACAACTGGGTACCTTC 779
QY 721 TGTAAAGCAATCCAGGAACCTACAAATACTAACAAAGCAACAAAGCAAACTAACAGTCA 780
Db 780 TGTAAAGCAATCCAGGAACCTACAAATACTAACAAAGCAACAAAGCAAACTAACAGTCA 839
QY 781 AGCAAGTCAAGTAATGACATGATAGTCTTCTTGAACAGCTCTCAAACTGCTTTGAG 840
Db 840 AGCAAGTCAAGTAATGACATGATAGTCTTCTTGAACAGCTCTCAAACTGCTTTGAG 899
QY 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGACAAATCACAGTCCGAA 900
Db 900 TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGACAAATCACAGTCCGAA 959
QY 901 AGCTAGAGTGTTCAGTGCCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAA 960
Db 960 AGCTAGAGTGTTCAGTGCCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAA 1019
QY 961 GTCTGAATTTGGAAGACGATCGCTCGTATATTCCTTCTGTTATCGTTTCAAACTGAG 1020
Db 1020 GTCTGAATTTGGAAGACGATCGCTCGTATATTCCTTCTGTTATCGTTTCAAACTGAG 1079
QY 1021 GGTACCAAGATTCAGGCGCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080
Db 1080 GGTACCAAGATTCAGGCGCAGAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1139
QY 1081 CCGGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTGTAGTCAAGTGT 1140
Db 1140 CCGGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTGTAGTCAAGTGT 1199
QY 1141 AGCAAGTGTGGGAGGATATGTTTCAAGAAAGGCGCTCTCTGTTATGCTTTTGC 1200
Db 1200 AGCAAGTGTGGGAGGATATGTTTCAAGAAAGGCGCTCTCTGTTATGCTTTTGC 1259
QY 1201 GAAAGATTTTACCCTGTTAACTGTTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 1260
Db 1260 GAAAGATTTTACCCTGTTAACTGTTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 1319
QY 1261 TGTTCACACACTTAACTGCTTAAATAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
Db 1320 TGTTCACACACTTAACTGCTTAAATAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1379
QY 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGCTCG 1380
Db 1380 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGCTCG 1439
QY 1381 TAAATCTGATTTCCCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGATCGACTAA 1440

```

QY	1021	GGTACGATTCAGGCCAGAACCAACCAAGTCCACCAAGCTCCGGAACCTAGTCCAGG	1080
DB		GGTACGATTCAGGCCAGAACCAACCAAGTCCACCAAGCTCCGGAACCTAGTCCAGG	1139
QY	1081	CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT	1140
DB		CCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGTGGT	1199
QY	1141	ACGAAAGCTGGGGAAGGATATGATTCGAAGAAAGGCACTCTCTCGTTATGTCTTTGC	1200
DB		ACGAAAGCTGGGGAAGGATATGATTCGAAGAAAGGCACTCTCTCGTTATGTCTTTGC	1259
QY	1201	GAAAGATTTACCATCTGAAACTGTGTTAAATAATCTTGAAGCAAGTTATCAAAAAGAGAG	1260
DB		GAAAGATTTACCATCTGAAACTGTGTTAAATAATCTTGAAGCAAGTTATCAAAAAGAGAG	1319
QY	1261	TGTTTCAACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGCTGACCAAGATTTTA	1320
DB		TGTTTCAACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGCTGACCAAGATTTTA	1379
QY	1321	TGATAAAGCATATAATCTGTTAACTGAGCTCATAAAGCCTTGTGTTGNAATAAAGGCTG	1380
DB		TGATAAAGCATATAATCTGTTAACTGAGCTCATAAAGCCTTGTGTTGNAATAAAGGCTG	1439
QY	1381	TAAATCTGATTTCCAGGCTTAGACAAATATTAGAACGCTTGAATGATGAAATCGACTAA	1440
DB		TAAATCTGATTTCCAGGCTTAGACAAATATTAGAACGCTTGAATGATGAAATCGACTAA	1499
QY	1441	TAAAGAAATTTGGTAGATGATTTATGTCATCTCTAGCACCCTATACCCATCCAGAGG	1500
DB		TAAAGAAATTTGGTAGATGATTTATGTCATCTCTAGCACCCTATACCCATCCAGAGG	1559
QY	1501	ACTTCGCAAAACCAATCTCTCAAAATGAGTATATCTGAAGACGAGTTCGTTATGCTCAAT	1560
DB		ACTTCGCAAAACCAATCTCTCAAAATGAGTATATCTGAAGACGAGTTCGTTATGCTCAAT	1619
QY	1561	AGCTGATAGTATACAAAGCTCAGATGGTTTACATTTTGTGATGAACATGATATATAGTGA	1620
DB		AGCTGATAGTATACAAAGCTCAGATGGTTTACATTTTGTGATGAACATGATATATAGTGA	1679
QY	1621	TGAAGAGATGATATGTAACGCTCATATGGCCATAGTCTGATGATTTGGAAGAGATAG	1680
DB		TGAAGAGATGATATGTAACGCTCATATGGCCATAGTCTGATGATTTGGAAGAGATAG	1739
QY	1681	CTTTCTGATAGGAAAGTTGTCAGCTCAAGCTTATCTAAAGAAAGAGTATCTTACC	1740
DB		CTTTCTGATAGGAAAGTTGTCAGCTCAAGCTTATCTAAAGAAAGAGTATCTTACC	1799
QY	1741	TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGGAGAGCTATTTA	1800
DB		TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGGAGAGCTATTTA	1859
QY	1801	CAATCGTGTGAAGGGGAAAGCAATTCCTCGTTCCGATTCCTATATATGTTGAGCA	1860
DB		CAATCGTGTGAAGGGGAAAGCAATTCCTCGTTCCGATTCCTATATATGTTGAGCA	1919
QY	1861	TACAGTTGAGTTTAAAGACGTTAAATTTGATTTATCTCTATAGGATCATTTACCATATAT	1920
DB		TACAGTTGAGTTTAAAGACGTTAAATTTGATTTATCTCTATAGGATCATTTACCATATAT	1979
QY	1921	TAAATTTGCTGTTGATGATCACAACAAGCTCCAAATGGCTATACCTTGAAGA	1980
DB		TAAATTTGCTGTTGATGATCACAACAAGCTCCAAATGGCTATACCTTGAAGA	2039
QY	1981	TTTGTGTTGCGATTAAGTACTACCTAGAACCTCTGACGAGCTCCACATTTCTAATGA	2040
DB		TTTGTGTTGCGATTAAGTACTACCTAGAACCTCTGACGAGCTCCACATTTCTAATGA	2099
QY	2041	TGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA	2100
DB		TGGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA	2159
QY	2101	TAAAGACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160

QY	1	TTCTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAATAATCTGTTTCTCTA	60
DB		TTCTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAATAATCTGTTTCTCTA	119
QY	61	TATAGATGGAATAACGAGCGCAAAACCGAGAAATTTGACTCTCTGATGAGTTAGCAA	120
DB		TATAGATGGAATAACGAGCGCAAAACCGAGAAATTTGACTCTCTGATGAGTTAGCAA	179
QY	121	CGGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACGAGGCTATGTCAC	180
DB		CGGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACGAGGCTATGTCAC	239
QY	181	TTCCATGCGGACCACTATCATTTATCAATGTTAAGGTTTCCTTATGACGCTATCATCAG	240
DB		TTCCATGCGGACCACTATCATTTATCAATGTTAAGGTTTCCTTATGACGCTATCATCAG	299
QY	241	TGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGGATATTGTTAATGA	300
DB		TGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGGATATTGTTAATGA	359
QY	301	GGTGAAGGTTGATATGTTATCAAGCTAGATGGAATACTATGTTTACCTTAAAGGATGC	360
DB		GGTGAAGGTTGATATGTTATCAAGCTAGATGGAATACTATGTTTACCTTAAAGGATGC	419
QY	361	TGCCACGCGGATAACGTCGCTACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG	420
DB		TGCCACGCGGATAACGTCGCTACAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG	479
QY	421	TCAACATCGTGAAGTGAACCTCCAGAAACGATGGTGTGTTGCTTGGGACGTTTCGCA	480
DB		TCAACATCGTGAAGTGAACCTCCAGAAACGATGGTGTGTTGCTTGGGACGTTTCGCA	539
QY	481	AGGAGCTTACTACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGATAC	540
DB		AGGAGCTTACTACAGATGATGTTTATCTTTAATGCTTCTGATATCATAGAGATAC	599
QY	541	TGCTGATGCTTATCTGTTCTCTCATGAGATCATTTACCATTTACATTTCTTAAAGAAATCGATT	600
DB		TGCTGATGCTTATCTGTTCTCTCATGAGATCATTTACCATTTACATTTCTTAAAGAAATCGATT	659
QY	601	ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCTTCTTATCTGTTGCGAGAAATCTGTCAA	660
DB		ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCTTCTTATCTGTTGCGAGAAATCTGTCAA	719
QY	661	TTCAAGAACTTATCGCGCAAAATAGCGATACACTTCAAGAACAACTGGGTACCTTC	720
DB		TTCAAGAACTTATCGCGCAAAATAGCGATACACTTCAAGAACAACTGGGTACCTTC	779
QY	721	TGTAAAGCAATCCAGGAACTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCA	780
DB		TGTAAAGCAATCCAGGAACTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCA	839
QY	781	AGCAAGTGAAGTAATGATGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAG	840
DB		AGCAAGTGAAGTAATGATGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAG	899
QY	841	TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTTCGAAC	900
DB		TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTTCGAAC	959
QY	901	AGCTAGAGTGTGCGATGCGCACAGGAGATCATTTACCATTCCTTCTTCTCTCAAAAT	960
DB		AGCTAGAGTGTGCGATGCGCACAGGAGATCATTTACCATTCCTTCTTCTCTCAAAAT	1019
QY	961	GTCCTGAATTTGGAAGAACGAATCGCTCGTATTATTCCTTCTGTTATCGTTCAAAACCATTC	1020
DB		GTCCTGAATTTGGAAGAACGAATCGCTCGTATTATTCCTTCTGTTATCGTTCAAAACCATTC	1079

QY 661 TTCAAGAACCTTATCGCGACAAAATAGGATATAACACATTCRAAGAACAACTCGGTACCTTC 720  
 Db 720 TTCAAGAACCTTATCGCGACAAAATAGGATATAACACATTCRAAGAACAACTCGGTACCTTC 779  
 QY 721 TGTAAAGCAATCCAGGAATCAAAATCTAACTAACAAGCAACCAACAGCAACACTAACAGTCA 780  
 Db 780 TGTAAAGCAATCCAGGAATCAAAATCTAACTAACAAGCAACCAACAGCAACACTAACAGTCA 839  
 QY 781 AGCAAGTCAAAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAAAGTCCCTTTGAG 840  
 Db 840 AGCAAGTCAAAAGTAATGACATGATAGTCTCTTGAACAGCTCTACAAAGTCCCTTTGAG 899  
 QY 841 TCAACGACATGTAGAAATCTGATGCGCTTGTCTTTGATCCACACAAATCAAAAGTCCGAAC 900  
 Db 900 TCAACGACATGTAGAAATCTGATGCGCTTGTCTTTGATCCACACAAATCAAAAGTCCGAAC 959  
 QY 901 AGCTAGAGTGTTCAGTGGCCACACAGGAGATCAATACCACTTCATCCCTTACTCTCAAAAT 960  
 Db 960 AGCTAGAGTGTTCAGTGGCCACACAGGAGATCAATACCACTTCATCCCTTACTCTCAAAAT 1019  
 QY 961 GTCTGAATTTGGAAGCAAGTCCGCTCGTATTAATTCCTGCTTGGTTATCGTTCAAAACCAATG 1020  
 Db 1020 GTCTGAATTTGGAAGCAAGTCCGCTCGTATTAATTCCTGCTTGGTTATCGTTCAAAACCAATG 1079  
 QY 1021 GGTACCAAGATTCAGGCGCAGAAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080  
 Db 1080 GGTACCAAGATTCAGGCGCAGAAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1139  
 QY 1081 CCGCAACCTTCAGCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGCTGGT 1140  
 Db 1140 CCGCAACCTTCAGCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGCTGGT 1199  
 QY 1141 ACAGAAAGTTGGGGAAGGATATGATTCGAAGCAAGGATCTCTGTTATGCTTTGTC 1200  
 Db 1200 ACAGAAAGTTGGGGAAGGATATGATTCGAAGCAAGGATCTCTGTTATGCTTTGTC 1259  
 QY 1201 GAAAGATTTACCACTCTGAAACTGTGTAAGCAAGTATCAAAACCAAGAGAG 1260  
 Db 1260 GAAAGATTTACCACTCTGAAACTGTGTAAGCAAGTATCAAAACCAAGAGAG 1319  
 QY 1261 TGTTTCACACATTTAACTGCTGTAAGCAAGTATGTTGCTCGTGACCAAGAAATTTTA 1320  
 Db 1320 TGTTTCACACATTTAACTGCTGTAAGCAAGTATGTTGCTCGTGACCAAGAAATTTTA 1379  
 QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAGAGCCTTGTGTTGNAATTAAGGTCG 1380  
 Db 1380 TGATAAGCATATAATCTGTTAACTGAGGCTCATAGAGCCTTGTGTTGNAATTAAGGTCG 1439  
 QY 1381 TAATTCGATTTCCAGCCTTAGACAAATTAATTAGAACGCTTGAATGATGAATCGACTAA 1440  
 Db 1440 TAATTCGATTTCCAGCCTTAGACAAATTAATTAGAACGCTTGAATGATGAATCGACTAA 1499  
 QY 1441 TAAAGAAAATTTGGTAGATGATTTATGGAATTCCTAGACCAATTAACCATCCAGAGCG 1500  
 Db 1500 TAAAGAAAATTTGGTAGATGATTTATGGAATTCCTAGACCAATTAACCATCCAGAGCG 1559  
 QY 1501 ACTTGGCAACCAAAATCTCAAAATTCAGTATCTGAAAGCAAGTTCGTTATTCCTCAAT 1560  
 Db 1560 ACTTGGCAACCAAAATCTCAAAATTCAGTATCTGAAAGCAAGTTCGTTATTCCTCAAT 1619  
 QY 1561 AGCTGATAAGTATACAAAGTTCAGATGGTTACATTTTGTGTAACATGATATATCAAGTGA 1620  
 Db 1620 AGCTGATAAGTATACAAAGTTCAGATGGTTACATTTTGTGTAACATGATATATCAAGTGA 1679  
 QY 1621 TGAAGGAGATGCATATGTAAGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1680  
 Db 1680 TGAAGGAGATGCATATGTAAGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG 1739  
 QY 1681 CCTTCTGTGTAAGGAAAAGTTGCAAGTCAAGCCTATCTAAGAAAAGGATCTCTACC 1740  
 Db 1740 CCTTCTGTGTAAGGAAAAGTTGCAAGTCAAGCCTATCTAAGAAAAGGATCTCTACC 1799

QY 1741 TCCATCTCCAGACCGCAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800  
 Db 1800 TCCATCTCCAGACCGCAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1859  
 QY 1801 CAATCGTGTGAAAGGGGAAAACGAATTCACACTCGTTCCGACTTCCATATATGTTTGAGCA 1860  
 Db 1860 CAATCGTGTGAAAGGGGAAAACGAATTCACACTCGTTCCGACTTCCATATATGTTTGAGCA 1919  
 QY 1861 TACAGTTGAGGTTAAAAACCGTAATTTGATTTATTCCTCATAAGGATCATTAACCAATATAT 1920  
 Db 1920 TACAGTTGAGGTTAAAAACCGTAATTTGATTTATTCCTCATAAGGATCATTAACCAATATAT 1979  
 QY 1921 TAAATTTGCTTGGTTTTCATGATCACATACAAAGCTCCAAAATGGCTATACCTTGGGAAGA 1980  
 Db 1980 TAAATTTGCTTGGTTTTCATGATCACATACAAAGCTCCAAAATGGCTATACCTTGGGAAGA 2039  
 QY 1981 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAAAGCTCCACATTTCTAATGA 2040  
 Db 2040 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAAAGCTCCACATTTCTAATGA 2099  
 QY 2041 TGGATGGGGCAATGCGCAGTGAAGTGTGTTAGGCAAGAAAGACCCACAGTGAAGATCCAAA 2100  
 Db 2100 TGGATGGGGCAATGCGCAGTGAAGTGTGTTAGGCAAGAAAGACCCACAGTGAAGATCCAAA 2159  
 QY 2101 TAAGAACTTCAAAGCGGATGAAGGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
 Db 2160 TAAGAACTTCAAAGCGGATGAAGGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2219  
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGACGAAAGTTTGTCTTGGGAA 2220  
 Db 2220 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGACGAAAGTTTGTCTTGGGAA 2279  
 QY 2221 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280  
 Db 2280 AGTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2339  
 QY 2281 TTTGACTCTTCAAATATGATTAACATAGTATCATGCGCAGAGCAGAAAATTAATCTTGC 2340  
 Db 2340 TTTGACTCTTCAAATATGATTAACATAGTATCATGCGCAGAGCAGAAAATTAATCTTGC 2399  
 QY 2341 GTTGTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2389  
 Db 2400 GTTGTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAAGGAAAAATAAAC 2448

RESULT 3  
 US-10-412-850-9  
 ; Sequence 9, Application US/10412850  
 ; Publication No. US20040001836A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-686  
 ; CURRENT APPLICATION NUMBER: US/10/412,850  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 9  
 ; LENGTH: 2451  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(2451)  
 ; OTHER INFORMATION: n = a, c, t or g  
 US-10-412-850-9



```

1381 TAAATCTGATTTCCAGCCTTAGACAAATATTAGACGCTTGAATGATGAATCGACTAA 1440
Db TAAATCTGATTTCCAGCCTTAGACAAATATTAGACGCTTGAATGATGAATCGACTAA 1440
1441 TAAAGAAAAATGGTAGATGATTTATGGCAATTCCTAGCACCAATACCCATCCAGAGG 1500
Db TAAAGAAAAATGGTAGATGATTTATGGCAATTCCTAGCACCAATACCCATCCAGAGG 1500
1501 ACTTGGCAAAACCAATTCCTAAATTCCTAAATTCCTAAATTCCTAAATTCCTAAAT 1560
Db ACTTGGCAAAACCAATTCCTAAATTCCTAAATTCCTAAATTCCTAAATTCCTAAAT 1560
1561 AGCTCATAGATATCAACGTCAGATGTTTACATTTTGTATGAACATGATATATCAGTGA 1620
Db AGCTCATAGATATCAACGTCAGATGTTTACATTTTGTATGAACATGATATATCAGTGA 1620
1621 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCTGATTTGAAAAAGATAG 1680
Db TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCTGATTTGAAAAAGATAG 1680
1681 CTTTCTGATAGAAAAAGTTGAGCTCAAGCCTATACCTAAGAAAAAGGTATCCTACC 1740
Db CTTTCTGATAGAAAAAGTTGAGCTCAAGCCTATACCTAAGAAAAAGGTATCCTACC 1740
1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTTA 1800
Db TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTTA 1800
1801 CAATCGTGTGAAAGGGAAAAAGCAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 1860
Db CAATCGTGTGAAAGGGAAAAAGCAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 1860
1861 TACAGTTGAGTTAAAAAGGTTAATTTGATTTATTCCTCATAGGATCATACCAATATAT 1920
Db TACAGTTGAGTTAAAAAGGTTAATTTGATTTATTCCTCATAGGATCATACCAATATAT 1920
1921 TAAATTTGCTGTTTGTATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
Db TAAATTTGCTGTTTGTATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1980
1981 TTTGTTTGGCAGATTAAGTACTACGTAGAACACCCCTGACGAACTCCATCTTAATGA 2040
Db TTTGTTTGGCAGATTAAGTACTACGTAGAACACCCCTGACGAACTCCATCTTAATGA 2040
2041 TGGATGGGCAATGCGAGTACGTAGTGTAGTGAACACCCCTGACGAACTCCATCTTAATGA 2100
Db TGGATGGGCAATGCGAGTACGTAGTGTAGTGAACACCCCTGACGAACTCCATCTTAATGA 2100
2101 TAAAGACTTCAAGCGGATGAAGGAGCAGTGAAGGAAACACCTGCTGAGCGAAGTCCC 2160
Db TAAAGACTTCAAGCGGATGAAGGAGCAGTGAAGGAAACACCTGCTGAGCGAAGTCCC 2160
2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAATCAAGAGACGAAAGTTTGTTCGGA 2220
Db TCAAGTAGAGACTGAAAAAGTAGAAGCCCAATCAAGAGACGAAAGTTTGTTCGGA 2220
2221 AGTAAAGGATTCAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
Db AGTAAAGGATTCAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280
2281 TTTGACTCTTCAAAATATGATTAACATAGTATATGCGAAGAGAGAAAAATTAATTCG 2340
Db TTTGACTCTTCAAAATATGATTAACATAGTATATGCGAAGAGAGAAAAATTAATTCG 2340
2341 GTTGTAAAAAGGAAGTAATCCTTCACTGTAAGTAAAGGAAAAAATAAAC 2389
Db GTTGTAAAAAGGAAGTAATCCTTCACTGTAAGTAAAGGAAAAAATAAAC 2389

```

RESULT 2  
 US-10-412-862-9  
 ; Sequence 9, Application US/10412862

```

; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John B.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; NAME/KEY: misc feature
; LOCATION: (1)-(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-412-862-9

Query Match 100.0%; Score 2389; DB 13; Length 2451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATATCGTGTTCCTA 60
Db TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATATCGTGTTCCTA 119
QY 61 TATAGATGAAAAACAGCGACCCAAAAACGGAGATTTGACTCCTCGATGAGGTTAGCAA 120
Db TATAGATGAAAAACAGCGACCCAAAAACGGAGATTTGACTCCTCGATGAGGTTAGCAA 179
QY 121 GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180
Db GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 239
QY 181 TTCACATGGCGACCACTATCATTTATCAATGGTAGGTTTCTTATGACGCTATCATCAG 240
Db TTCACATGGCGACCACTATCATTTATCAATGGTAGGTTTCTTATGACGCTATCATCAG 299
QY 241 TGAAGATTTACTCATCAAGATCCAACTATAAGCTAAAAGATGAGGATTTGTTAATGA 300
Db TGAAGATTTACTCATCAAGATCCAACTATAAGCTAAAAGATGAGGATTTGTTAATGA 359
QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 419
QY 361 TGGCCACGGCGATTAACGTCGTTACAAAGAGGAAATCAATCGACAAAAACAAGGATAG 420
Db TGGCCACGGCGATTAACGTCGTTACAAAGAGGAAATCAATCGACAAAAACAAGGATAG 479
QY 421 TCACCATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTGTTGGCTTGGCAGCGTTGCGCA 480
Db TCACCATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTGTTGGCTTGGCAGCGTTGCGCA 539
QY 481 AGGACGCTATCTACATACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
Db AGGACGCTATCTACATACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGATAC 599
QY 541 TGGTGTATGCTTATATGTTTCTCATGGAGATCATACCATTAATGCTTCTGATATCATAGAGATG 600
Db TGGTGTATGCTTATATGTTTCTCATGGAGATCATACCATTAATGCTTCTGATATCATAGAGATG 659
QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCTTATCTGCTGCGAGGAAATCTGTCAA 660
Db ATCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCTTATCTGCTGCGAGGAAATCTGTCAA 719

```



88 19 0.8 7493 15 US-10-187-319-5 GENERAL INFORMA  
89 19 0.8 7493 15 US-10-131-510A-5 Sequence 5, Appl 1  
90 19 0.8 91278 17 US-10-322-281-341 Sequence 341, App

## ALIGNMENTS

```

RESULT 1
US-09-765-272-55
; Sequence 55, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-765-272-55

Query Match      100.0%; Score 2389; DB 9; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TTCTTACGAGTGGGAGTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA 60
Db      1  TTCTTACGAGTGGGAGTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA 60

QY      61  TATAGATGGAATCAAGCGACGCAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
Db      61  TATAGATGGAATCAAGCGACGCAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA 120

QY      121  GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 180
Db      121  GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 180

QY      181  TTCATATGGCGACCACTATCATTTACAAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db      181  TTCATATGGCGACCACTATCATTTACAAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240

QY      241  TGAAGATTAATCATGTAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300

```

```

Db      241  TGAAGATTAATCATGTAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300
QY      301  GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
Db      301  GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGC 360
QY      361  TGCCACGCGGATAACGTCGTACAAAAGAGGAATCAATCGACAAAACAAAGAGCATAG 420
Db      361  TGCCACGCGGATAACGTCGTACAAAAGAGGAATCAATCGACAAAACAAAGAGCATAG 420
QY      421  TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTGCTTGCCTTGGCAGCTTCGCA 480
Db      421  TCAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTGCTTGCCTTGGCAGCTTCGCA 480
QY      481  AGGACGCTATATACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db      481  AGGACGCTATATACAGATGATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
QY      541  TGGTGATGCTTATATCGTTCCTCATGGAGATCAATACCATTAATTCCTAAGATGAGTT 600
Db      541  TGGTGATGCTTATATCGTTCCTCATGGAGATCAATACCATTAATTCCTAAGATGAGTT 600
QY      601  ATCAGCTAGCGAGTTGGCTCTCGAGAAAGCTTCTATCTGCTCGAGAAATCTGTCAA 660
Db      601  ATCAGCTAGCGAGTTGGCTCTCGAGAAAGCTTCTATCTGCTCGAGAAATCTGTCAA 660
QY      661  TTCAAGAACCTATCGCGGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
Db      661  TTCAAGAACCTATCGCGGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
QY      721  TGTAAAGCAATCCAGGAACTAATACTAACAAGCAACAAAGCAACAACTAAGTCA 780
Db      721  TGTAAAGCAATCCAGGAACTAATACTAACAAGCAACAAAGCAACAACTAAGTCA 780
QY      781  AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAACCTGCTTGG 840
Db      781  AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAACCTGCTTGG 840
QY      841  TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCACAAATCAAGTCA 900
Db      841  TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCACAAATCAAGTCA 900
QY      901  AGCTAGAGGTGTTGCGAGTCCACACGAGATCAATACCACTTCCCTTACTCTCAAT 960
Db      901  AGCTAGAGGTGTTGCGAGTCCACACGAGATCAATACCACTTCCCTTACTCTCAAT 960
QY      961  GTCTGAATTTGGAAGAACGAATCGCTCGTATTAATTCCTTTCCTTTCGTTTCAACCATTC 1020
Db      961  GTCTGAATTTGGAAGAACGAATCGCTCGTATTAATTCCTTTCCTTTCGTTTCAACCATTC 1020
QY      1021  GGTACCAGATTCAAGCCAGAACCAAGTCCCAACCGAGCTCCGGAACCTAGTCCAG 1080
Db      1021  GGTACCAGATTCAAGCCAGAACCAAGTCCCAACCGAGCTCCGGAACCTAGTCCAG 1080
QY      1081  CCGCAACCTGCAACAAATCTTAAATAGACTCAAAATCTCTTCTTGTGTTAGTCACTGGT 1140
Db      1081  CCGCAACCTGCAACAAATCTTAAATAGACTCAAAATCTCTTCTTGTGTTAGTCACTGGT 1140
QY      1141  ACGAAAGTTGGGAAAGGATATGTTTTCGAAGAAAAGGCGATCTCTCGTTATGCTTTTC 1200
Db      1141  ACGAAAGTTGGGAAAGGATATGTTTTCGAAGAAAAGGCGATCTCTCGTTATGCTTTTC 1200
QY      1201  GAAAGATTTTACCATCTGAACCTGTTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 1260
Db      1201  GAAAGATTTTACCATCTGAACCTGTTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 1260
QY      1261  TGTTTCACACACTTTTAACTGCTTAAAGAAATGTTGCTTCTCGTGAACAAGATTTTA 1320
Db      1261  TGTTTCACACACTTTTAACTGCTTAAAGAAATGTTGCTTCTCGTGAACAAGATTTTA 1320
QY      1321  TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAGGGTCG 1380
Db      1321  TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAGGGTCG 1380

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 01:51:39 ; Search time 1910 Seconds  
(without alignments)  
6340.940 Million cell updates/sec

Title: US-09-765-271-55  
Perfect score: 2389  
Sequence: 1 TTCTACGAGTGGACTGT.....TAAGTAAGGAAATAAATAC 2389

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3340653 seqs, 2534783454 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2389	100.0	2389	9	US-09-765-272-55
2	2389	100.0	2451	13	US-10-412-862-9
3	2389	100.0	2451	16	US-10-412-850-9
4	2389	100.0	2451	16	US-10-387-783-9
5	2338	97.9	8195	13	US-10-158-844-94
6	148	6.2	2523	10	US-09-884-465A-3
7	148	6.2	2647	10	US-09-884-465A-4
8	53	2.2	2290	9	US-09-765-272-65
9	53	2.2	2359	13	US-10-158-844-243
10	53	2.2	2481	10	US-09-769-787-206
11	53	2.2	2531	13	US-10-412-862-5
12	53	2.2	2531	13	US-10-412-862-11
13	53	2.2	2531	16	US-10-412-850-5
14	53	2.2	2531	16	US-10-412-850-11

53	15	2.2	2531	16	US-10-387-783-5	Sequence 5, Appli
53	16	2.2	2531	16	US-10-387-783-11	Sequence 11, Appli
53	17	2.2	2639	10	US-09-884-465A-5	Sequence 5, Appli
44	18	1.8	973	13	US-10-158-844-355	Sequence 355, App
19	30	1.3	40	9	US-09-765-272-282	Sequence 282, App
20	30	1.3	40	13	US-10-412-862-3	Sequence 3, Appli
21	30	1.3	40	16	US-10-412-850-3	Sequence 3, Appli
22	30	1.3	40	16	US-10-387-783-3	Sequence 3, Appli
23	27	1.1	36	13	US-10-412-862-1	Sequence 1, Appli
24	27	1.1	36	16	US-10-412-850-1	Sequence 1, Appli
25	27	1.1	36	16	US-10-387-783-1	Sequence 1, Appli
26	27	1.1	37	9	US-09-765-272-281	Sequence 281, App
27	26	1.1	35	13	US-10-412-862-2	Sequence 2, Appli
28	26	1.1	35	16	US-10-412-850-2	Sequence 2, Appli
29	26	1.1	35	16	US-10-387-783-2	Sequence 2, Appli
30	25	1.0	33	10	US-09-884-465A-45	Sequence 45, Appli
31	25	1.0	34	10	US-09-884-465A-47	Sequence 47, Appli
32	25	1.0	35	10	US-09-884-465A-174	Sequence 174, App
33	24	1.0	1342	9	US-09-765-272-181	Sequence 181, App
34	24	1.0	1455	10	US-09-769-787-246	Sequence 246, App
35	24	1.0	1455	13	US-09-769-744A-23	Sequence 23, Appli
36	24	1.0	1455	16	US-10-412-862-7	Sequence 7, Appli
37	24	1.0	1455	16	US-10-412-850-7	Sequence 7, Appli
38	24	1.0	1455	16	US-10-387-783-7	Sequence 7, Appli
39	24	1.0	2528	10	US-09-884-465A-9	Sequence 9, Appli
40	24	1.0	3120	10	US-09-884-465A-1	Sequence 1, Appli
41	24	1.0	5048	10	US-09-884-465A-2	Sequence 2, Appli
42	24	1.0	6867	13	US-10-158-844-192	Sequence 192, App
43	22	0.9	228139	13	US-10-087-192-232	Sequence 232, App
44	21	0.9	29	10	US-09-884-465A-44	Sequence 44, Appli
45	20	0.8	501	16	US-10-275-933-3	Sequence 3, Appli
46	20	0.8	1146	10	US-09-769-736-23	Sequence 23, Appli
47	20	0.8	2469	10	US-09-769-736-17	Sequence 17, Appli
48	20	0.8	5158	16	US-10-275-933-11	Sequence 11, Appli
49	20	0.8	5215	10	US-09-252-088-13	Sequence 13, Appli
50	20	0.8	5215	16	US-10-340-792-13	Sequence 13, Appli
51	19	0.8	164	9	US-09-983-965-4749	Sequence 4749, Ap
52	19	0.8	246	17	US-10-437-963-1529	Sequence 1529, Ap
53	19	0.8	427	11	US-09-801-944B-62	Sequence 62, Appli
54	19	0.8	483	17	US-10-021-323-6488	Sequence 6488, Ap
55	19	0.8	509	17	US-10-021-323-6205	Sequence 6205, Ap
56	19	0.8	526	17	US-10-021-323-12910	Sequence 12910, A
57	19	0.8	533	13	US-10-027-632-231121	Sequence 231121,
58	19	0.8	533	13	US-10-027-632-231122	Sequence 231122,
59	19	0.8	533	13	US-10-027-632-231123	Sequence 231123,
60	19	0.8	533	16	US-10-027-632-231121	Sequence 231121,
61	19	0.8	533	16	US-10-027-632-231122	Sequence 231122,
62	19	0.8	533	16	US-10-027-632-231123	Sequence 231123,
63	19	0.8	561	17	US-10-021-323-1518	Sequence 1518, Ap
64	19	0.8	565	17	US-10-021-323-10473	Sequence 10473, A
65	19	0.8	593	17	US-10-021-323-12821	Sequence 12821, A
66	19	0.8	594	17	US-10-021-323-9719	Sequence 9719, Ap
67	19	0.8	599	17	US-10-021-323-7454	Sequence 7454, Ap
68	19	0.8	601	17	US-10-021-323-8160	Sequence 8160, Ap
69	19	0.8	658	17	US-10-767-701-25410	Sequence 25410, A
70	19	0.8	678	13	US-10-424-599-66507	Sequence 66507, A
71	19	0.8	693	16	US-10-012-697-610	Sequence 610, App
72	19	0.8	984	17	US-10-437-963-6825	Sequence 6825, Ap
73	19	0.8	1020	13	US-10-282-122A-16763	Sequence 16763, A
74	19	0.8	1112	13	US-10-027-632-118539	Sequence 118539,
75	19	0.8	1112	13	US-10-027-632-118540	Sequence 118540,
76	19	0.8	1112	16	US-10-027-632-118539	Sequence 118539,
77	19	0.8	1112	16	US-10-027-632-118540	Sequence 118540,
78	19	0.8	1112	17	US-10-437-963-9121	Sequence 9121, Ap
79	19	0.8	1155	17	US-10-437-963-87931	Sequence 87931, A
80	19	0.8	1213	13	US-10-027-632-117222	Sequence 117222,
81	19	0.8	1213	13	US-10-027-632-117223	Sequence 117223,
82	19	0.8	1213	13	US-10-027-632-117224	Sequence 117224,
83	19	0.8	1213	16	US-10-027-632-117222	Sequence 117222,
84	19	0.8	1213	16	US-10-027-632-117223	Sequence 117223,
85	19	0.8	1213	16	US-10-027-632-117224	Sequence 117224,
86	19	0.8	1239	17	US-10-767-795-794	Sequence 794, App
87	19	0.8	1455	10	US-09-769-736-71	Sequence 71, Appli

```

US-08-474-503-5
; Sequence 5, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 1..407
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "5'UTR"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 7471..7476
; OTHER INFORMATION: /function= "PolyA_signal"
; FEATURE:
; NAME/KEY: repeat unit
; LOCATION: 7368..7493
; OTHER INFORMATION: /rpt_type= "terminal"
; OTHER INFORMATION: /note= "3'UTR"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 408..7367
; OTHER INFORMATION: /product= "Coagulation Factor VIII"
; PUBLICATION INFORMATION:
; AUTHORS: Elder, F.
; AUTHORS: Lakich, D.
; AUTHORS: Gitschier, J.
; TITLE: Sequence of the Murine Factor VIII cDNA.
; Patent No. 5744446
; JOURNAL: Genomics
; VOLUME: 16
; PAGES: 374-379
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
US-08-474-503-5

```

Query Match

0.8%; Score 19; DB 1; Length 7493;

Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2347 AAAAGGAAGTAATCCTTCA 2365  
| | | | | | | | | | | | | | | |  
Db 206 AAAAGGAAGTAATCCTTCA 224

Search completed: October 1, 2004, 03:35:56  
Job time : 175 secs

```
Sequence 895, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebeschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 895:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1559RP
US-08-998-416-895

Query Match 0.8%; Score 19; DB 3; Length 622;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 TCAAGAACAACTGGGTAC 716
Db 383 TCAAGAACAACTGGGTAC 401

RESULT 24
US-08-212-133A-7
Sequence 7, Application US/08212133A
Patent No. 5663060
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Runge, Marschall S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia

Country: US
ZIP: 30303
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "5'UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7471..7476
OTHER INFORMATION: /function= "PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "3'UTR"
FEATURE:
NAME/KEY: misc feature
LOCATION: 408..7367
OTHER INFORMATION: /product= "Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 5663060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

Query Match 0.8%; Score 19; DB 1; Length 7493;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2347 AAAAGGAAGTAATCCTTCA 2365
Db 206 AAAAGGAAGTAATCCTTCA 224

RESULT 25
```

```

; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340Pl
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-961-527-192
;
; Query Match 1.0%; Score 24; DB 4; Leng
; Best Local Similarity 100.0%; Pred.No.0.039;
; Matches 24; Conservative 0; Mismatches 0;
;
; QY 546 ATGCTTATATCGTTCCTCATGGAG 569
; |||||||
; DB 6190 ATGCTTATATCGTTCCTCATGGAG 6167
;
; RESULT 22
; US-09-489-039A-6492
; Sequence 6492, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQU
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND T
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 6492
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
;
; US-09-489-039A-6492
;
; Query Match 0.8%; Score 20; DB 4; Leng
; Best Local Similarity 100.0%; Pred.No.4;
; Matches 20; Conservative 0; Mismatches 0;
;
; QY 2368 TGTAAAGTAAGAAAAATAA 2387
; |||||||
; DB 183 TGTAAAGTAAGAAAAATAA 202
;
; RESULT 23
; US-08-998-416-895

```

STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 281:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 281:  
US-09-536-784-281

Query Match 1.1%; Score 27; DB 4; Length 37;  
Best Local Similarity 100.0%; Pred. No. 0.00093; Indels 0;  
Matches 27; Conservative 0; Mismatches 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGC 27  
DB 11 TTCTTACGAGTTGGGACTGTATCAAGC 37

RESULT 17  
US-09-468-656A-2  
Sequence 2, Application US/09468656A  
Patent No. 6582706  
GENERAL INFORMATION:  
APPLICANT: Johnson, Leslie S.  
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural  
TITLE OF INVENTION: Motifs  
FILE REFERENCE: 469201-444  
CURRENT APPLICATION NUMBER: US/09/468,656A  
CURRENT FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: 60/113,048  
PRIOR FILING DATE: 1998-12-21  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 35  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Forward primer  
OTHER INFORMATION: used in amplification of the Sp36 gene sequence.  
US-09-468-656A-2

Query Match 1.1%; Score 26; DB 4; Length 35;  
Best Local Similarity 100.0%; Pred. No. 0.003;  
Matches 26; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;

QY 16 ACTGTATCAAGCTAGAACGGTTAAGG 41  
DB 10 ACTGTATCAAGCTAGAACGGTTAAGG 35

RESULT 18  
US-08-961-083-181  
Sequence 181, Application US/08961083  
Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 181:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1342 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-083-181

Query Match 1.0%; Score 24; DB 3; Length 1342;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCTCTCATGGAG 569  
DB 525 ATGCTTATATCGTTCTCTCATGGAG 548

RESULT 19  
US-09-536-784-181  
Sequence 181, Application US/09536784  
Patent No. 6573082  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

```

;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 282:
US-09-536-784-282

Query Match      1.3%; Score 30; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2360 CTTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 40 CTTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 13
US-09-468-656A-3/c
; Sequence 3, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1998-12-048
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-3

Query Match      1.3%; Score 30; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2360 CTTTCATCTGTAAGTAAGGAAAAATAAAC 2389
Db 40 CTTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 14
US-09-468-656A-1
; Sequence 1, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR FILING DATE: 1998-12-048
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 36
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Forward primer
; OTHER INFORMATION: used in amplification of the Sp36 gene sequence.
US-09-468-656A-1
```

```

Query Match      1.1%; Score 27; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
Db 10 TTCTTACGAGTTGGGACTGTATCAAGC 36

RESULT 15
US-08-961-083-281
; Sequence 281, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 281:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-281

Query Match      1.1%; Score 27; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTTACGAGTTGGGACTGTATCAAGC 27
Db 11 TTCTTACGAGTTGGGACTGTATCAAGC 37

RESULT 16
US-09-536-784-281
; Sequence 281, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
```

```
QY 983 GCTGTTATTTCCCTTCGTTTCATGTTCAACCAATTCGGTACCAAGATTCAAG 1035
      |||||||
Db 1000 GCTGTTATTTCCCTTCGTTTCATGTTCAACCAATTCGGTACCAAGATTCAAG 1052

RESULT 10
US-08-961-527-355
; Sequence 355, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 355:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-355

Query Match 1.8%; Score 44; DB 4; Length 973;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 CATTACCATTCATTCTTAAGATGATGTTATTCAGTAGCGAGTT 615
      |||||||
Db 722 CATTACCATTCATTCTTAAGATGATGTTATTCAGTAGCGAGTT 765

RESULT 11
US-08-961-083-282/c
; Sequence 282, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
```

```
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-282

Query Match 1.3%; Score 30; DB 3; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2360 CCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
      |||||||
Db 40 CCTTCATCTGTAAGTAAGGAAAAATAAAC 11

RESULT 12
US-09-536-784-282/c
; Sequence 282, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```



```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-243

Query Match      2.2%; Score 53; DB 4; Length 2359;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    983  GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAGATTCCAAG 1035
       |||||||
Db     1879  GTCGTAATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAGATTCCAAG 1931

RESULT 8
US-09-468-656A-5
; Sequence 5, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-5

Query Match      2.2%; Score 53; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY    983  GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAGATTCCAAG 1035
       |||||||
Db     1003  GCTCGTATTATTCCTCCCTTCGTTATCGTTCAAAACCATTTGGGTACCAGATTCCAAG 1055

RESULT 9
US-09-468-656A-11
; Sequence 11, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-11

Query Match      2.2%; Score 53; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

4253 GAAAGATTTCACCTCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 4312  
1261 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTGTAACCAAGAAATTTTA 1320  
4313 TGTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCCTGTAACCAAGAAATTTTA 4372  
1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTGNAATTAAGGGTCG 1380  
4373 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTGNAATTAAGGGTCG 4432  
1391 TAAATTCATATTCAGGCTTAGACAAATATTAGAACGCTTGAATGATGATGATGATGATGAT 1440  
4433 TAAATTCATATTCAGGCTTAGACAAATATTAGAACGCTTGAATGATGATGATGATGATGAT 4492  
1441 TAAAGAAAAATGTTAGATGATTTATTTGCAATCTTAGACCAATATCCCATCCAGAGCG 1500  
4493 TAAAGAAAAATGTTAGATGATTTATTTGCAATCTTAGACCAATATCCCATCCAGAGCG 4552  
1501 ACTTGCGCAAAACCAATTCCTCAATTTAGATGATGATGATGATGATGATGATGATGATGAT 1560  
4553 ACTTGCGCAAAACCAATTCCTCAATTTAGATGATGATGATGATGATGATGATGATGATGAT 4612  
1561 AGCTGATAGTATACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
4613 AGCTGATAGTATACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4672  
1621 TGAAGGAGATGATATGTAACGCTCATATGAGGCTCATATGAGGCTCATATGAGGCTCATAT 1680  
4673 TGAAGGAGATGATATGTAACGCTCATATGAGGCTCATATGAGGCTCATATGAGGCTCATAT 4732  
1681 CTTTCTGATAGGAAAAAGTTGACGCTCAAGCTTATCTAAAGAAAGGTTATCTTACC 1740  
4733 CTTTCTGATAGGAAAAAGTTGACGCTCAAGCTTATCTAAAGAAAGGTTATCTTACC 4792  
1741 TCCATCTCAGAGCGAGATGTTAAAGCAATCACTGAGATGATGAGGCTCATATGAGGCTCATAT 1800  
4793 TCCATCTCAGAGCGAGATGTTAAAGCAATCACTGAGATGATGAGGCTCATATGAGGCTCATAT 4852  
1801 CAACTGCTGTAAGAGGGAAGAAACGAATTCACCTGCTGATGATGATGATGATGATGATGATGAT 1860  
4853 CAACTGCTGTAAGAGGGAAGAAACGAATTCACCTGCTGATGATGATGATGATGATGATGATGAT 4912  
1861 TACAGTTGAGTTTAAAGGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
4913 TACAGTTGAGTTTAAAGGCTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 4972  
1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
4973 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5032  
1981 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTGAAGCAAGCTTCAATTTCAATGA 2040  
5033 TTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTGAAGCAAGCTTCAATTTCAATGA 5092  
2041 TGGATGGGGCAATGCCAGTGAAGTATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100  
5093 TGGATGGGGCAATGCCAGTGAAGTATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 5152  
2101 TAAAGCACTTCAAGCGGATGAAGGACGAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
5153 TAAAGCACTTCAAGCGGATGAAGGACGAGTGAAGGAAACACCTGCTGAGCCAGAGTCCC 5212  
2161 TCAAGTGAAGACTGAAAGGATGAAGGACCACTCAAGAGCAAGAGTTTGTCTTGGAA 2220  
5213 TCAAGTGAAGACTGAAAGGATGAAGGACCACTCAAGAGCAAGAGTTTGTCTTGGAA 5272  
2221 AGTAACGATTTCTAGTCTGAAGGCAATGCAACAGAACTCTAGTCTGTTTACGAATAA 2280  
5273 AGTAACGATTTCTAGTCTGAAGGCAATGCAACAGAACTCTAGTCTGTTTACGAATAA 5332  
2281 TTTGACTCTTCAATTTATGGAATAACAATAGTATCATGCGAGAGCAGAGAAATTAATCTTGC 2340  
5333 TTTGACTCTTCAATTTATGGAATAACAATAGTATCATGCGAGAGCAGAGAAATTAATCTTGC 5392

2341 GTTGTAAAGGAAGTAATCTTCAATCTGTAAGTAAGGAAAAATAAAC 2389  
5393 GTTGTAAAGGAAGTAATCTTCAATCTGTAAGTAAGGAAAAATAAAC 5441

RESULT 5  
US-08-961-083-65  
; Sequence 65, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2290 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-961-083-65

Query Match 2.2%; Score 53; DB 3; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 5.4e-17;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

983 GCTCGTATTATTCCTCTGTTTCAAAACCATTTGGTACCAGATTCAG 1035  
944 GCTCGTATTATTCCTCTGTTTCAAAACCATTTGGTACCAGATTCAG 996

RESULT 6  
US-09-536-784-65  
; Sequence 65, Application US/09536784  
; Patent No. 6573082  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:



		Best Local Similarity 100.0%; Pred. No. 0;		Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACCGGTTAAGGAAATAATATCGTGTTCCTA	60		
Db	60	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACCGGTTAAGGAAATAATATCGTGTTCCTA	119		
Qy	61	TATAGATCGAAAAAAGGAGCGCAAAAGGAGAAATTTGACTCTGATGAGGTTAGCAA	120		
Db	120	TATAGATCGAAAAAAGGAGCGCAAAAGGAGAAATTTGACTCTGATGAGGTTAGCAA	179		
Qy	121	CGGTGAAGGATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC	180		
Db	180	CGGTGAAGGATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC	239		
Qy	181	TTCAATCGGACCACTATCAATTTACAATGGTAAGGTTCTTATGAGCGCTATCATCAG	240		
Db	240	TTCAATCGGACCACTATCAATTTACAATGGTAAGGTTCTTATGAGCGCTATCATCAG	299		
Qy	241	TGAAGAAATTACTCATGAAGATCCAAACTATAAGCTTAAGATGAGGATTTGTTAATGA	300		
Db	300	TGAAGAAATTACTCATGAAGATCCAAACTATAAGCTTAAGATGAGGATTTGTTAATGA	359		
Qy	301	GGTCAAGGTTGATATGTTATCAAGGTAGATCGAAATACTATGTTTACCTTAAAGGATGC	360		
Db	360	GGTCAAGGTTGATATGTTATCAAGGTAGATCGAAATACTATGTTTACCTTAAAGGATGC	419		
Qy	361	TGCCCAACGCGGATFAACGTCCTGTAACAAAGAGGAAATCAATCGACAAAAACAAGAGATAG	420		
Db	420	TGCCCAACGCGGATFAACGTCCTGTAACAAAGAGGAAATCAATCGACAAAAACAAGAGATAG	479		
Qy	421	TCAAATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGCTTGGCAGGTTCCGCA	480		
Db	480	TCAAATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGCTTGGCAGGTTCCGCA	539		
Qy	481	AGGACGCTACTACAGATGATGGTTATCTTTAATGCTTCTGATATCATAGAGGATAC	540		
Db	540	AGGACGCTACTACAGATGATGGTTATCTTTAATGCTTCTGATATCATAGAGGATAC	599		
Qy	541	TGTTGATGCTTATATCTCTCATGAGATCAATACCAATTAATTCCTTAAGATGAGTT	600		
Db	600	TGTTGATGCTTATATCTCTCATGAGATCAATACCAATTAATTCCTTAAGATGAGTT	659		
Qy	601	ATCAGCTAGGAGTTGGCTGCTCAGAGACCTTCTATCTGTCGAGGAATCTGTCAA	660		
Db	660	ATCAGCTAGGAGTTGGCTGCTCAGAGACCTTCTATCTGTCGAGGAATCTGTCAA	719		
Qy	661	TTCAAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC	720		
Db	720	TTCAAGAACTTATCGCGCAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTC	779		
Qy	721	TGTAAGCAATCCAGGAATCAAAATCTAAACAAGCAACAAGCAACACTAACAGTCA	780		
Db	780	TGTAAGCAATCCAGGAATCAAAATCTAAACAAGCAACAAGCAACACTAACAGTCA	839		
Qy	781	AGCAAGTCAAGTAATGACATTTGATGCTCTTTGAACAGCTCTACAAACCTGCTTTGAG	840		
Db	840	AGCAAGTCAAGTAATGACATTTGATGCTCTTTGAACAGCTCTACAAACCTGCTTTGAG	899		
Qy	841	TCAACGATGATGAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTCAAGC	900		
Db	900	TCAACGATGATGAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTCAAGC	959		
Qy	901	AGCTAGAGGTTGTCAGTGCACACGAGGATCAATACCACTTATCGCTTACTCTCAAT	960		
Db	960	AGCTAGAGGTTGTCAGTGCACACGAGGATCAATACCACTTATCGCTTACTCTCAAT	1019		
Qy	961	GTCTGAATTCGAAGCAAGATCGCTCTGATTTATTTCCCTTCTGTTATCGTTCAAACCATG	1020		
Db	1020	GTCTGAATTCGAAGCAAGATCGCTCTGATTTATTTCCCTTCTGTTATCGTTCAAACCATG	1079		
Qy	1021	GGTACCAGATTCAGGCGCAGAACCAACAGTCCCAACCGACTCCGGAACTAGTCCAGG	1080		

Db	1080	GGTACCAGATTCAGGCGCAGAACCAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG	1139
Qy	1081	CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAATTTCTTCTTTGGTTAGTCACTGGT	1140
Db	1140	CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAATTTCTTCTTTGGTTAGTCACTGGT	1199
Qy	1141	ACGAAAAGTTGGGGAAGGATATGATTCGAGAAAAGGCGATCTCTCGTTATGTTCTTGC	1200
Db	1200	ACGAAAAGTTGGGGAAGGATATGATTCGAGAAAAGGCGATCTCTCGTTATGTTCTTGC	1259
Qy	1201	GAAAGATTTACCATCTGAACTGTTAAATACTTGAAGAGCAAGTTTATCAAAAACAAGAGAG	1260
Db	1260	GAAAGATTTACCATCTGAACTGTTAAATACTTGAAGAGCAAGTTTATCAAAAACAAGAGAG	1319
Qy	1261	TGTTTTCAACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTGACCAAGAAATTTTA	1320
Db	1320	TGTTTTCAACACACTTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTGACCAAGAAATTTTA	1379
Qy	1321	TGATAAAGCATATAATCTGTTAACTGAGGCTCATTAAGCCCTGTTTGNAAATAAGGCTCG	1380
Db	1380	TGATAAAGCATATAATCTGTTAACTGAGGCTCATTAAGCCCTGTTTGNAAATAAGGCTCG	1439
Qy	1381	TAATTTCTGANTTTCCAGCCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA	1440
Db	1440	TAATTTCTGANTTTCCAGCCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA	1499
Qy	1441	TAAAGAAAAATTTGGTAGATGATTTATTGTCATTTCTTAGCACCAATTAACCATCCAGAGCG	1500
Db	1500	TAAAGAAAAATTTGGTAGATGATTTATTGTCATTTCTTAGCACCAATTAACCATCCAGAGCG	1559
Qy	1501	ACTTGGCAAAACCAAAATTTCAAATTTGAGTATACTGAAAGCGAAGTTTCGTATTCCTCAAT	1560
Db	1560	ACTTGGCAAAACCAAAATTTCAAATTTGAGTATACTGAAAGCGAAGTTTCGTATTCCTCAAT	1619
Qy	1561	AGCTGATAGTATACAAAGCTCAGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA	1620
Db	1620	AGCTGATAGTATACAAAGCTCAGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA	1679
Qy	1621	TGAAGAGATGTCATATGTAACGCTCATATGCGCCATAGTCTACTGGATTTGGAAGAAATAG	1680
Db	1680	TGAAGAGATGTCATATGTAACGCTCATATGCGCCATAGTCTACTGGATTTGGAAGAAATAG	1739
Qy	1681	CTTTTCTGATTAAGGAAAAAGTTGAGCTCAAGCTTATCTAAAGAAAAGGATATCTTACC	1740
Db	1740	CTTTTCTGATTAAGGAAAAAGTTGAGCTCAAGCTTATCTAAAGAAAAGGATATCTTACC	1799
Qy	1741	TCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAAGATAGTGCAGCAGCTATTTA	1800
Db	1800	TCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAAGATAGTGCAGCAGCTATTTA	1859
Qy	1801	CAATCGTGTGAAAGGGGAAAAAGGATTTCCACTCGTTTCGACTTCCATATATGTTGAGCA	1860
Db	1860	CAATCGTGTGAAAGGGGAAAAAGGATTTCCACTCGTTTCGACTTCCATATATGTTGAGCA	1919
Qy	1861	TACAGTTGAGGTTAAAAACCGGTAATTTGATTTATTCCTCATAGGATCAITTACCATAATAT	1920
Db	1920	TACAGTTGAGGTTAAAAACCGGTAATTTGATTTATTCCTCATAGGATCAITTACCATAATAT	1979
Qy	1921	TAAATTTGCTTGGTTTGTGATGATCACATACAAAGCTCCAAATGGCTATATACCTTGGAGA	1980
Db	1980	TAAATTTGCTTGGTTTGTGATGATCACATACAAAGCTCCAAATGGCTATATACCTTGGAGA	2039
Qy	1981	TTTGTGTTGCGACGATTAAGTACTACGTPAGAACACCTTCGACGAACGCTCCACATTTCAATGA	2040
Db	2040	TTTGTGTTGCGACGATTAAGTACTACGTPAGAACACCTTCGACGAACGCTCCACATTTCAATGA	2099
Qy	2041	TGGATGGGCAATGCCAGTGAAGCTGTTAGGCAAGAAAGACCAACAGTGAAGATCCAAA	2100
Db	2100	TGGATGGGCAATGCCAGTGAAGCTGTTAGGCAAGAAAGACCAACAGTGAAGATCCAAA	2159
Qy	2101	TAAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	2160	TAAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2219



1561 AGCTGATAGTATACAACGTCAGATGGTTACATTTTGGATGAATATATCAATCAGTGA 1620  
1621 TGAAGGAGATGATATGTAAGCCCTCATATGGGCCATAGTCTACTGGATTGGAAGATAG 1680  
1621 TGAAGGAGATGATATGTAAGCCCTCATATGGGCCATAGTCTACTGGATTGGAAGATAG 1680  
1681 CCTTCTCATAGGAAAAGTTGCGAGCTCAAGCCTATATCTAAGAAAAGGTTATCCTACC 1740  
1681 CCTTCTCATAGGAAAAGTTGCGAGCTCAAGCCTATATCTAAGAAAAGGTTATCCTACC 1740  
1741 TCCATCTCCAGACGCGAGATGTTTAAAGCAAATCCAACTCGAGATAGTGCAGCAGCTATTTA 1800  
1741 TCCATCTCCAGACGCGAGATGTTTAAAGCAAATCCAACTCGAGATAGTGCAGCAGCTATTTA 1800  
1801 CAATCGTGTGAAGGGGAAAAGAAAGAAATTCCTCGTTCGATTCATATATGTTGAGCA 1860  
1801 CAATCGTGTGAAGGGGAAAAGAAAGAAATTCCTCGTTCGATTCATATATGTTGAGCA 1860  
1861 TACAGTTGAGGTTTAAAGCGGTAATTTGATTATTCCTCATAGGATCATTTACCAATAT 1920  
1861 TACAGTTGAGGTTTAAAGCGGTAATTTGATTATTCCTCATAGGATCATTTACCAATAT 1920  
1921 TAAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980  
1921 TAAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGA 1980  
1981 TTTGTTGGGAGATTAAGTACTAGTAGAACAACCTGACAGCCGTAAGCTTCAATCTAATGA 2040  
1981 TTTGTTGGGAGATTAAGTACTAGTAGAACAACCTGACAGCCGTAAGCTTCAATCTAATGA 2040  
2041 TGGATGGGCAATGCGAGTACATGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100  
2041 TGGATGGGCAATGCGAGTACATGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100  
2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
2161 TCAAGTAGAGATGAAAAGTAGAAGCCCACTCAAGAGCAGAGTTTTCCTGCGAA 2220  
2161 TCAAGTAGAGATGAAAAGTAGAAGCCCACTCAAGAGCAGAGTTTTCCTGCGAA 2220  
2221 AGTAACGGATCTAGTCTCAAGGCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280  
2221 AGTAACGGATCTAGTCTCAAGGCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280  
2281 TTTGACTCTTCAAAATTTATGATAAACAATAGTATCGGAGAGCAAGAAATTTACTTGC 2340  
2281 TTTGACTCTTCAAAATTTATGATAAACAATAGTATCGGAGAGCAAGAAATTTACTTGC 2340  
2341 GTTGTAAAGGAAGTAATCCTCTCATCTGTAAGTAGGAAAATAAAC 2389  
2341 GTTGTAAAGGAAGTAATCCTCTCATCTGTAAGTAGGAAAATAAAC 2389

RESULT 2  
US-09-536-784-55  
; Sequence 55, Application US/09536784  
; Patent No. 6573082  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2389 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 55:  
US-09-536-784-55  
Query Match 100.0%; Score 2389; DB 4; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60  
Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTA 60  
QY 61 TATAGATGAAAACAAAGCGACGCAAAACGAGAAATTTGACTTCCTGATGAGGTTAGCAA 120  
Db 61 TATAGATGAAAACAAAGCGACGCAAAACGAGAAATTTGACTTCCTGATGAGGTTAGCAA 120  
QY 121 GCGTGAAGGAATCAATCGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180  
Db 121 GCGTGAAGGAATCAATCGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180  
QY 181 TTCAATGCGGACCACTATCATATTATTAATGGTAAGGTTCTTATGAGCCTATCATCAG 240  
Db 181 TTCAATGCGGACCACTATCATATTATTAATGGTAAGGTTCTTATGAGCCTATCATCAG 240  
QY 241 TGAAGAAATTTACTCATCAAGATCCAAACTATAAGCTAAGGATGAGGATTTGTTAATGA 300  
Db 241 TGAAGAAATTTACTCATCAAGATCCAAACTATAAGCTAAGGATGAGGATTTGTTAATGA 300  
QY 301 GGTCAAGGTTGGATATGTTTCAAGGTAGATGAAAATACTATGTTTACCTTAAGGATGC 360  
Db 301 GGTCAAGGTTGGATATGTTTCAAGGTAGATGAAAATACTATGTTTACCTTAAGGATGC 360  
QY 361 TGCCCAACGCGGATTAACGTCGCTACAAAAGAGAAATCAATCGCAAAAAACAAGAGCATAG 420  
Db 361 TGCCCAACGCGGATTAACGTCGCTACAAAAGAGAAATCAATCGCAAAAAACAAGAGCATAG 420  
QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTTGCCTTGCACGTTTCGCA 480  
Db 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTTGCCTTGCACGTTTCGCA 480  
QY 481 AGGACGCTATCTACAGATGATGGTTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540  
Db 481 AGGACGCTATCTACAGATGATGGTTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540  
QY 541 TGGTATGCTTATATCGTTCCTCATCGAGATCATTTACCAATTTACATTCTTAAGAAATGAGTT 600  
Db 541 TGGTATGCTTATATCGTTCCTCATCGAGATCATTTACCAATTTACATTCTTAAGAAATGAGTT 600  
QY 601 ATCAGCTAGGAGTTGCTGCTGTCAGAAAGCTTCTTATCTGCTGAGGAAATCTGTCAAA 660  
Db 601 ATCAGCTAGGAGTTGCTGCTGTCAGAAAGCTTCTTATCTGCTGAGGAAATCTGTCAAA 660

APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESS: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2389 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-083-55

Query Match 100.0%; Score 2389; DB 3; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TTCTTAGAGTTGGAGCTGATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA	60
DB	1	TTCTTAGAGTTGGAGCTGATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA	60
QY	61	TATAGATGGAAACAAAGCGACGCAAAAACGGAGAAATTCACCTCCTGATGAGGTAGCAA	120
DB	61	TATAGATGGAAACAAAGCGACGCAAAAACGGAGAAATTCACCTCCTGATGAGGTAGCAA	120
QY	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC	180
DB	121	GCCTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC	180
QY	181	TTCAATGGCGACCACTATCAATATTAACAATGTTAGGTTCCCTTATGACCTATCATCAG	240
DB	181	TTCAATGGCGACCACTATCAATATTAACAATGTTAGGTTCCCTTATGACCTATCATCAG	240
QY	241	TGAAGAAATTAATCAATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAATGA	300
DB	241	TGAAGAAATTAATCAATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTGTTAATGA	300
QY	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC	360
DB	301	GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC	360
QY	361	TGCCACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACCAAGAGCATAG	420
DB	361	TGCCACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACCAAGAGCATAG	420
QY	421	TCACATCTGTAAGGTGGAACTCAAGAAACGATGGTGTGCTGTTGCCCTGCGACGTTGCGA	480
DB	421	TCAACATCTGTAAGGTGGAACTCAAGAAACGATGGTGTGCTGTTGCCCTGCGACGTTGCGA	480

QY	481	AGACGCTATACACATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	540
DB	481	AGACGCTATACACATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	540
QY	541	TGGTATGCTTATATCTGCTCATGAGATCATACCATTTACATTTCTTAAGATGAGTT	600
DB	541	TGGTATGCTTATATCTGCTCATGAGATCATACCATTTACATTTCTTAAGATGAGTT	600
QY	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTATCTGGTCGAGAAATCTGTCAA	660
DB	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTATCTGGTCGAGAAATCTGTCAA	660
QY	661	TTCAAGAACTATCGCGCAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTC	720
DB	661	TTCAAGAACTATCGCGCAAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTC	720
QY	721	TGTAAGCAATCCAGGAATCAAAATCTAACAAGCAACAAAGCAACAACTACACAGTCA	780
DB	721	TGTAAGCAATCCAGGAATCAAAATCTAACAAGCAACAAAGCAACAACTACACAGTCA	780
QY	781	AGCAAGTCAAAAGTAAATGACATTCATAGTCTCTTGAACAGCTCTACAACTGCTTTGAG	840
DB	781	AGCAAGTCAAAAGTAAATGACATTCATAGTCTCTTGAACAGCTCTACAACTGCTTTGAG	840
QY	841	TCAACGACATGTAGAAATCTGATGCGCTTGTCTTGTATCCAGCAAAATCAACAGTCA	900
DB	841	TCAACGACATGTAGAAATCTGATGCGCTTGTCTTGTATCCAGCAAAATCAACAGTCA	900
QY	901	AGCTAGAGTGTTCAGTGCACACGAGATCATTTACCACTTCTCTTCTTCTTCTTCTTCTT	960
DB	901	AGCTAGAGTGTTCAGTGCACACGAGATCATTTACCACTTCTCTTCTTCTTCTTCTTCTT	960
QY	961	GTCTGAAATTTGGAAGAAAGCAATCGTCTGATTTATTCCTCTTCTTCTTCTTCTTCTT	1020
DB	961	GTCTGAAATTTGGAAGAAAGCAATCGTCTGATTTATTCCTCTTCTTCTTCTTCTTCTT	1020
QY	1021	GGTACCAGATTTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACTTACCTCAGG	1080
DB	1021	GGTACCAGATTTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACTTACCTCAGG	1080
QY	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTCTTCTTCTTCTTCTT	1140
DB	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAAAATCTTCTTCTTCTTCTTCTTCTT	1140
QY	1141	ACGAAAGTTGGGAAAGGATATGTTATTCGAAGAAAGGAGGATCTCTCTGTTTCTT	1200
DB	1141	ACGAAAGTTGGGAAAGGATATGTTATTCGAAGAAAGGAGGATCTCTCTGTTTCTT	1200
QY	1201	GAAAGATTTTACCATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG	1260
DB	1201	GAAAGATTTTACCATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG	1260
QY	1261	TGTTTCCACACCTTTAACTGCTTAAAGAAAGGATGTTGCTCTCTGTTTCTTCTTCTT	1320
DB	1261	TGTTTCCACACCTTTAACTGCTTAAAGAAAGGATGTTGCTCTCTGTTTCTTCTTCTT	1320
QY	1321	TGTAAGCAATTAATCTGTTTAACTGAGGCTCATTAAGGCTTGTGTTGTTGTTGTTG	1380
DB	1321	TGTAAGCAATTAATCTGTTTAACTGAGGCTCATTAAGGCTTGTGTTGTTGTTGTTG	1380
QY	1381	TAAATCTGATTTTCAAGGCTTACACAAATTTATAGAACGCTTGAATGATGAACTAA	1440
DB	1381	TAAATCTGATTTTCAAGGCTTACACAAATTTATAGAACGCTTGAATGATGAACTAA	1440
QY	1441	TAAAGAAATTTGGTAGATGATTTTATTTGGCAATTTCTAGCACCATAATCCCATCAG	1500
DB	1441	TAAAGAAATTTGGTAGATGATTTTATTTGGCAATTTCTAGCACCATAATCCCATCAG	1500
QY	1501	ACTTGGCAACCAATTTCTCAAAATGAGTATCTGAAGACGAGTTCTGTTCTCAAT	1560
DB	1501	ACTTGGCAACCAATTTCTCAAAATGAGTATCTGAAGACGAGTTCTGTTCTCAAT	1560
QY	1561	AGCTGATAGTATACAACTGAGTGGTTTACATTTTGTGATGACATGATATATCATGTA	1620







```

source
1. .267
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0203J06"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUC2M library"
/notes="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 0.8%; Score 20; DB 28; Length 267;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1711 AGCCTACTAAGAAAAAG 1730
|||||
Db 71 AGCCTACTAAGAAAAAG 90

RESULT 25
BB803034
LOCUS
DEFINITION
BB803034 RIKEN full-length enriched, 16 days neonate male
dienecephalon Mus musculus cDNA clone G630051N24 3', mRNA sequence.
BB803034
VERSION
BB803034.1 GI:16975663
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 371)
REFERENCE
AUTHORS
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
JOURNAL
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/

```

```

FEATURES
            Location/Qualifiers
            source
            1..371
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="G630051N24"
            /sex="male"
            /tissue_type="dienecephalon"
            /dev_stage="16 days neonate"
            /clone_lib="RIKEN full-length enriched, 16 days neonate
            male dienecephalon"

ORIGIN
Query Match 0.8%; Score 20; DB 10; Length 371;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 TTTCCTATATAGATGGA AAA 73
|||||
Db 200 TTTCCTATATAGATGGA AAA 219

Search completed: October 1, 2004, 03:33:05
Job time : 6095 secs

```

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (10), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

[illegible]





reannealing hybridization was used."

```

ORIGIN
Query Match          0.9%; Score 21; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAAGCAGAGTTT 2210
      |||||
Db 359 AACTCAAGAAGCAGAGTTT 339

RESULT 16
BUI15421/c
LOCUS
DEFINITION
603140883F1 CSEQCHL15 Gallus gallus CDNA clone CHEST132622 5', mRNA
sequence.
ACCESSION
BUI15421
VERSION
BUI15421.1 GI:25321521
KEYWORDS
EST.
SOURCE
Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 754)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
PUBMED
12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 754
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST132622"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHL15"
/notes="Organ: brain; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent #
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BsgI and
BamHI sites [5'gccgcgtgcagcccgatcgcgaaaaag]
[5'aattcttttccggtccgggtgcgcgc]"

ORIGIN
Query Match          0.9%; Score 21; DB 13; Length 754;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAAGCAGAGTTT 2210
      |||||
Db 667 AACTCAAGAAGCAGAGTTT 647

RESULT 17
BUI211482
LOCUS
DEFINITION
604159635F1 CSEQCHN03 Gallus gallus CDNA clone CHEST1013e10 5',
mRNA sequence.
ACCESSION
BUI211482
VERSION
BUI211482.1 GI:25385632
KEYWORDS
EST.
SOURCE
Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 769)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
PUBMED
12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 769
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST1013e10"
/tissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN03"
/notes="Organ: whole embryo; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methyated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

```

```

ORIGIN
Query Match          0.9%; Score 21; DB 13; Length 769;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAAGCAGAGTTT 2210
      |||||
Db 389 AACTCAAGAAGCAGAGTTT 409

RESULT 18
BUI308927/c
LOCUS
DEFINITION
603538542F1 CSEQCHN60 Gallus gallus CDNA clone CHEST504c11 5', mRNA
sequence.
ACCESSION
BUI308927
VERSION
BUI308927.1 GI:25816928
KEYWORDS
EST.
SOURCE
Gallus gallus (chicken)

```

COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

FEATURES source  
1..690  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Haruna Nijo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="bags35c04"  
/tissue\_type="shoots"  
/dev\_stage="germination"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna  
Nijo germination shoots"

## ORIGIN

Query Match 0.9%; Score 21; DB 12; Length 690;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1078 AGGCCCGCAACCTGCACCAA 1098  
|||||  
DB 76 AGGCCCGCAACCTGCACCAA 96

RESULT 14  
BU403931/c  
LOCUS  
DEFINITION BU403931 731 bp mRNA linear EST 27-NOV-2002  
604141302F1 CSEQCHN59 Gallus gallus cDNA clone CHEST974d2 5', mRNA  
sequence.

ACCESSION BU403931  
VERSION BU403931.1 GI:25772987  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 731)  
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
TITLE A Comprehensive Collection of Chicken cDNAs  
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
MEDLINE 22335534  
PubMed 12445392

COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)

PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

FEATURES source  
1..731

/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST974d2"  
/dev\_stage="36"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN59"  
/note="Organ: limbs; Vector: pBluescript II KS(+); Site\_1:  
EcORI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.

## ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 731;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAAGTTT 2210  
|||||  
DB 531 AACTCAAGACGACGAAGTTT 511

RESULT 15  
BU283494/c

LOCUS  
DEFINITION BU283494 742 bp mRNA linear EST 27-NOV-2002  
603862178F1 CSEQCHN54 Gallus gallus cDNA clone CHEST875k20 5', mRNA  
sequence.

ACCESSION BU283494  
VERSION BU283494.1 GI:25732950  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 742)  
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
TITLE A Comprehensive Collection of Chicken cDNAs  
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
MEDLINE 22335534  
PubMed 12445392

COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)

PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.

FEATURES source  
1..742

/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton Line 151"  
/db\_xref="taxon:9031"  
/clone="CHEST875k20"  
/sex="Female"  
/tissue\_type="not cerebrum or cerebellum"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN54"  
/note="Organ: brain; Vector: pBluescript II KS(+); Site\_1:  
EcORI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcORI, size-selected, and cloned into the NotI and EcORI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6  
(1996) 791, except that a significantly longer

Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcORI, size-selected, and cloned into the NotI and EcORI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6  
(1996) 791, except that a significantly longer  
reannealing hybridization was used."

Query Match 0.9%; Score 21; DB 28; Length 640;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1539 ACGAAGTTCGATTGCTCAAT 1559  
 |||||  
 Db 332 ACGAAGTTCGATTGCTCAAT 312

RESULT 11  
 BU345704/c  
 LOCUS  
 DEFINITION 604170839F1 CSEQCHN68 Gallus gallus CDNA clone CHEST1033m10 5',  
 mRNA sequence.  
 ACCESSION BU345704  
 VERSION BU345704.1 GI:25853705  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Gallus gallus (chicken)  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 664)  
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
 source  
 1..664  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Compton Line 151"  
 /db\_xref="taxon:9031"  
 /clone="CHEST1033m10"  
 /sex="Female"  
 /tissue\_type="cerebellum"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN68"  
 /notes="Organ: brain; Vector: pBluescript II KS(+); Site\_1:  
 EcoRI; Site 2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

ORIGIN  
 Query Match 0.9%; Score 21; DB 13; Length 664;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAGCAGAGATT 2210  
 |||||  
 Db 295 AACTCAAGAGCAGAGATT 275

RESULT 12  
 AG084591/c  
 LOCUS  
 DEFINITION Pan troglodytes DNA, clone: PTB-082G04.F, genomic survey sequence.  
 ACCESSION AG084591  
 VERSION AG084591.1 GI:16636393  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes

REFERENCE 1  
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE BAC end sequences of Library PTB  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 681)  
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
 Totoki,Y., Watanabe,H. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 [E-mail:chimbesc@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170]  
 COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end  
 was generated during the R&D process and may have higher chance of  
 clone tracking errors.  
 PRIMERS

LIBRARY Sequencing: -21M13  
 Vector : pKS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI.  
 Location/Qualifiers  
 1..681  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-082G04.F"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"

FEATURES  
 source  
 1..681  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-082G04.F"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"

## ORIGIN

Query Match 0.9%; Score 21; DB 29; Length 681;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1755 CAGATGTTAAGCAATCCAA 1775  
 |||||  
 Db 142 CAGATGTTAAGCAATCCAA 122

RESULT 13  
 BU549886  
 LOCUS  
 DEFINITION

BU549886 690 bp mRNA linear EST 14-NOV-2002  
 BU549886 K. Sato unpublished cDNA library, cv. Haruna Niho  
 germination shoots Hordeum vulgare subsp. vulgare cDNA clone  
 bags35c04 3', mRNA sequence.

ACCESSION BU549886  
 VERSION BU549886.1 GI:24968336  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poaceae; fruticeae; Hordeum.  
 REFERENCE 1 (bases 1 to 690)  
 AUTHORS Sato,K., Saisho,D. and Takeda,K.  
 TITLE Barley EST sequencing project in NIG and Okayama Univ  
 JOURNAL Unpublished (2002)

/db\_xref="taxon:7159"  
 /clone="Notre Dame Liverpool-9714"  
 /clone\_lib="Notre Dame Liverpool"  
 /note="Vector: pECBAC1; Site\_1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"

## ORIGIN

Query Match 0.9%; Score 21; DB 29; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1222 TGTAAAAAATCTTGAAGCAA 1242  
 DB 27 TGTAAAAAATCTTGAAGCAA 7

RESULT 9  
 BU397983/c  
 LOCUS 581 bp mRNA linear EST 27-NOV-2002  
 DEFINITION 603534550F1 CSEQCHN58 Gallus gallus cDNA clone CHEST494c13 5', mRNA sequence.  
 ACCESSION BU397983  
 VERSION BU397983.1 GI:25767039  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 581)  
 AUTHORS Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

## FEATURES

Location/Qualifiers  
 1..581  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Compton Line 151"  
 /db\_xref="taxon:9031"  
 /clone="CHEST494c13"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN58"

/note="Organ: small intestine; Vector: pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

## ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 581;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGACGACGAAGTTT 2210  
 DB 368 AACTCAAGACGACGAAGTTT 348

RESULT 10  
 AZ391393/c  
 LOCUS 640 bp DNA linear GSS 03-OCT-2000  
 DEFINITION 1M0153A03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0153A03 R, genomic survey sequence.

ACCESSION AZ391393  
 VERSION AZ391393.1 GI:10506436  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 640)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0153 row: A column: 03  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 640.

## FEATURES

Location/Qualifiers  
 1..640  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0153A03"  
 /sex="Male"  
 /lab\_host="F. Coli strain XL10-Gold, TI-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN



```

RESULT 6
BI544663          499 bp      mRNA      linear      EST 05-SEP-2001
LOCUS             603242619F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5284997 5',
DEFINITION        mRNA sequence.
ACCESSION         BI544663
VERSION           BI544663.1 GI:15431975
KEYWORDS          EST.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE         1 (bases 1 to 499)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Robert Strausberg, Ph.D.
                  Email: cgabbs-rc@mail.nih.gov
                  Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
                  cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
                  Toshiyuki and Piero Carninci (RIKEN)
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLNL1720 row: f column: 06
                  High quality sequence stop: 299.
FEATURES          Location/Qualifiers
                  1..499
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:5284997"
                   /tissue_type="hippocampus"
                   /lab_host="DH10B"
                   /clone_lib="NIH_MGC_95"
                   /notes="Organ: brain; Vector: pBluescriptR (modified
                   pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                   gtcgag); Oligo-dT primed using primer
                   5'-TTTTTTTTTTTTTTVN-3', size-selected for average
                   insert size 2.5 kb and normalized to ROT 5. This is a
                   primary library enriched for full-length clones and
                   constructed using the Cap-trapper method (Carninci, in
                   preparation). Library constructed by M. Brownstein
                   (NIH/NHGRI, National Institutes of Health). Note: this
                   is a NIH_MGC Library."
ORIGIN
Query Match      0.9%; Score 21; DB 12; Length 499;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 GAGGAATCAATCGACAAAAA 409
      |||||||
Db 414 GAGGAATCAATCGACAAAAA 434

RESULT 7
AW932212
LOCUS             AW932212
DEFINITION        EST358055 tomato fruit mature green, TAMU Lycopersicon esculentum
ACCESSION         AW932212
VERSION           AW932212.1 GI:8107613
KEYWORDS          EST.
SOURCE            Lycopersicon esculentum (tomato)
ORGANISM          Lycopersicon esculentum
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE         1 (bases 1 to 549)

```

```

AUTHORS           Alcalá,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
                  Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ann,S.,
                  Ronning,C.M., Fraser,C.M., Martin,G.B., Tankeley,S.D. and
                  Giovannoni,J.
TITLE             Generation of ESTs from tomato fruit tissue
JOURNAL           Unpublished (1999)
COMMENT           Contact: CUGI
                  Clemson University Genomics Institute
                  Clemson University
                  100 Jordan Hall, Clemson, SC 29634, USA
                  Email: http://www.genome.clemson.edu/orders/index.html
                  5 prime sequence.
FEATURES          Location/Qualifiers
                  1..549
                   /organism="Lycopersicon esculentum"
                   /mol_type="mRNA"
                   /cultivar="TA496"
                   /db_xref="taxon:4081"
                   /clone="CLEF47L13"
                   /tissue_type="fruit pericarp"
                   /dev_stage="mature green (3-5 days pre-ripening)"
                   /lab_host="SOLR"
                   /clone_lib="tomato fruit mature green, TAMU"
                   /notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
                   XhoI; CLSF - Fruit were tagged at the 1cm stage and
                   harvested 3-5 days prior to ripening. Fruit were cut in
                   half to verify the seeds were indeed 'immature' and the
                   seeds and locules were discarded prior to freezing the
                   pericarp"
ORIGIN
Query Match      0.9%; Score 21; DB 10; Length 549;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AGGTGGATATGTTATCAAGG 326
      |||||||
Db 135 AGGTGGATATGTTATCAAGG 155

RESULT 8
CC847187/c
LOCUS             CC847187
DEFINITION        ND.L.9714.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
                  Notredame Liverpool-9714, genomic survey sequence.
ACCESSION         CC847187
VERSION           CC847187.1 GI:33194851
KEYWORDS          GSS.
SOURCE            Aedes aegypti (yellow fever mosquito)
ORGANISM          Aedes aegypti
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
                  Stegomyia.
REFERENCE         1 (bases 1 to 551)
AUTHORS           Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE             BAC end sequencing of Aedes aegypti
JOURNAL           Unpublished (2003)
COMMENT           Other_GSSs: ND.L.9714.T7
                  Contact: Brendan Loftus
                  Department of Eukaryotic Genomics
                  TIGR
                  9712 Medical Center Drive, Rockville, MD 20850, USA
                  Tel: 301-838-3543
                  Fax: 301-838-0208
                  Email: enta@tigr.org
                  Library was provided by David Severson
                  Seq primer: SP6
                  Class: BAC ends.
FEATURES          Location/Qualifiers
                  1..551
                   /organism="Aedes aegypti"
                   /mol_type="genomic DNA"
                   /strain="liverpool"

```

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda PLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

#### FEATURES

Location/Qualifiers  
1..618  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAPL15-41-A11"  
/tissue\_type="mixture of silique and flower"  
/lab\_host="DH10B"  
/clone\_lib="RAPL15"  
/note="Site\_1: BamHI; Site\_2: SalI"

#### ORIGIN

Query Match 0.9%; Score 22; DB 9; Length 618;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2285 ACTCTTCAAAATTATCGATAACA 2306

DB 481 ACTCTTCAAAATTATCGATAACA 502

#### RESULT 4

BH970231/c  
LOCUS odf81d11.b1 B.oleracea002 Brassica oleracea genomic, genomic survey  
DEFINITION sequence.  
ACCESSION BH970231  
VERSION BH970231.1 GI:23453048  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 682)

Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash,W., Rabinowicz,P.D. and Wilson,R.K.

Whole genome shotgun reads from Brassica oleracea

Unpublished (2002)

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

Plate: odf81 row: d column: 11

Seq primer: -21UpPOT forward

Class: shotgun

High quality sequence start: 14

High quality sequence stop: 519.

Location/Qualifiers

1..682

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/db\_xref="taxon:3712"

/clone\_lib="B.oleracea002"

/note="Vector: pORW13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

#### ORIGIN

Query Match 0.9%; Score 22; DB 28; Length 682;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1811 AAAGGGGAAAACGAATTCAC 1832

DB 44 AAAGGGGAAAACGAATTCAC 23

#### RESULT 5

BU473013/c  
LOCUS 603365019F1 CSEQRBN21 Gallus gallus cDNA clone ChEST262f5 5', mRNA  
DEFINITION sequence.  
ACCESSION BU473013  
VERSION BU473013.1 GI:25966590  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 411)

Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

22335534

12445332

COMMENT

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: [Simon.Hubbard@umist.ac.uk](mailto:Simon.Hubbard@umist.ac.uk).

#### FEATURES

source

1..411

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="Layer"

/db\_xref="taxon:9031"

/clone="ChEST262f5"

/sex="Female"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="CSEQRBN21"

/note="Organ: ovary; Vector: pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA

synthesis was initiated using an oligo(dT) primer, using

methylated C in the first strand synthesis reaction.

Following this first strand reaction, double-stranded cDNA

was bluntended, ligated to NotI adapters, digested with

EcoRI, size-selected, and cloned into the NotI and EcoRI

compatible sites of a custom modified MCS of the

pBluescript (KS+) vector. The library was normalized in 2

rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6

(1996): 791, except that a significantly longer

reannealing hybridization was used."

#### ORIGIN

Query Match 0.9%; Score 21; DB 13; Length 411;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2190 AACTCAAGAAGCAGAGTTT 2210

DB 273 AACTCAAGAAGCAGAGTTT 253

```

78      20      0.8      923      10      BG033498      602301896
79      20      0.8      924      29      CG862220
80      20      0.8      957      13      BQ712408
81      20      0.8      984      29      CG858044
82      20      0.8      2185     11      AK037932
83      19      0.8      66      28      BH907895
84      19      0.8      110     28      CC021347
85      19      0.8      136     14      H74951
86      19      0.8      158     29      CE567072
87      19      0.8      175     10      BE505229
88      19      0.8      186     28      BH867080
89      19      0.8      220      9      AV236691
90      19      0.8      250     14      CF765303

```

## ALIGNMENTS

```

RESULT 1
CE751062/c
LOCUS
DEFINITION
tigr-gss-dog-17000369611786 Dog Library Canis familiaris genomic,
genomic survey sequence.
CE751062
CE751062.1 GI:37091679
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
MEDLINE
14512627
PUBMED
COMMENT
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. .469
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

```

## FEATURES

```

source
1. .469
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

```

## ORIGIN

```

Query Match 1.0%; Score 23; DB 29; Length 469;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2140 ACCTGCTGAGCCAGAGTCCCTC 2162
|||||
Db 269 ACCTGCTGAGCCAGAGTCCCTC 247
|||||

```

## RESULT 2

```

BE528097
LOCUS
DEFINITION
M70N06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
Clone 600037123R1 5', mRNA sequence.
ACCESSION
BE528097
VERSION
BE528097.1 GI:9786075

```

```

KEYWORDS
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 261)
AUTHORS
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilardua, O., Jaworski, J.G., Ohlgrogge, J. and Benning, C.
TITLE
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL
Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE
20567808
PUBMED
11115876
COMMENT
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI
48824, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.

```

## FEATURES

```

source
1. .261
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600037123R1"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"

```

## ORIGIN

```

Query Match 0.9%; Score 22; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2285 ACTCTCAAATTTATGATAACA 2306
|||||
Db 200 ACTCTCAAATTTATGATAACA 221
|||||

```

## RESULT 3

```

AU236995
LOCUS
DEFINITION
AU236995 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-41-A11 5',
mRNA sequence.
ACCESSION
AU236995
VERSION
AU236995.1 GI:19876164

```

## KEYWORDS

```

SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 618)
AUTHORS
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J.,
Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A.,
Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 22:20:23 : Search time 6083 Seconds  
(without alignments)  
11727.889 Million cell updates/sec

Title: US-09-765-271-55  
Perfect score: 2389  
Sequence: 1 TTCTTACGAGTGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htr:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htr:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	1.0	469	29	CE751062 tigr-gss-
C 2	22	0.9	261	10	BE528097 M70N06STM
C 3	22	0.9	618	9	AU236995
C 4	22	0.9	682	28	BH970231 odff81d11.

BU473013	603365019	411	13	BU473013
BU544663	603242619	499	12	BU544663
AW932212	EST358055	549	10	AW932212
CC847187	NDL.9714.	551	29	CC847187
BU397983	603534550	581	13	BU397983
AZ391393	IM0153A03	640	28	AZ391393
BU345704	604170839	664	13	BU345704
AG084591	Pan trogl	681	29	AG084591
BU549886	BJ549886	690	12	BU549886
BU403931	604141302	731	13	BU403931
BU283494	603862178	742	13	BU283494
BU115421	603408863	754	13	BU115421
BU211482	604159635	769	13	BU211482
BU308927	603538542	799	13	BU308927
CC088871	CSU-K33r.	870	28	CC088871
BU273390	603531527	935	13	BU273390
BE538572	BB538572	175	10	BE538572
AZ502341	IM0341023	228	28	AZ502341
AZ769909	IM0571J01	233	28	AZ769909
AZ943045	2M023J06	267	28	AZ943045
BB803034	BB803034	371	10	BB803034
CG109167	PUKDL80TB	401	29	CG109167
AQ046037	RPC111-36	406	28	AQ046037
AQ216823	HS-3253_B	471	28	AQ216823
BZ685217	PUBCP45TD	496	28	BZ685217
AW239782	ptiic.pk0	540	10	AW239782
CK229737	702668904	558	14	CK229737
CF256901	pha004_c0	582	14	CF256901
BH369198	AG-ND-139	584	28	BH369198
CE778613	tigr-gss-	584	29	CE778613
BH766941	BMBAC345E	588	28	BH766941
CK229735	702482441	589	14	CK229735
CA370410	650577_NC	595	13	CA370410
BW282910	BW282910	603	13	BW282910
CK229736	702568991	607	14	CK229736
AZ524614	234PDD06	622	28	AZ524614
AQ12913	RPC1-11-1	630	28	AQ12913
BH765160	BMBAC354A	636	28	BH765160
BW112810	BW112810	643	13	BW112810
BW243771	BW243771	649	13	BW243771
BG598682	602703065	651	12	BG598682
BF296373	036PDB02	656	10	BF296373
BH085526	RPC1-24-2	656	28	BH085526
AZ373971	IM0126P17	658	28	AZ373971
BW261677	BW261677	664	13	BW261677
BG911162	602806563	668	12	BG911162
CE140831	tigr-gss-	676	29	CE140831
AJ453442	AJ453442	689	9	AJ453442
BW262397	BW262397	689	13	BW262397
BF294340	004PBA08	698	10	BF294340
BB527573	BB527573	706	10	BB527573
BW027074	BW027074	708	13	BW027074
CK229738	702668904	708	14	CK229738
AV862261	AV862261	712	9	AV862261
CA763329	BR060002B	713	14	CA763329
CG851011	ZMMBBB033	725	29	CG851011
AL141617	Anopheles	736	29	AN01PFB4
CG098656	PUIWE15TD	746	29	CG098656
BW132597	BW132597	748	13	BW132597
BZ995097	PUDPB72TD	761	28	BZ995097
BW131850	BW131850	765	13	BW131850
AL150861	Anopheles	770	29	CNS01MFG
CB598673	AGENCOURT	776	14	CB598673
BE130716	L48-922T3	792	10	BE130716
BH374713	AG-ND-161	795	28	BH374713
BU566263	BU566263	809	12	BU566263
CC159651	IG03C07.9	812	28	CC159651
CC577803	CH240_456	818	29	CC577803
BH045505	RPCI-24-3	831	28	BH045505
AL440694	T3 end of	876	29	CNS07DNK
CF255654	mdvnl30_e	881	14	CF255654
CG109169	PUKDL80TD	896	29	CG109169
CF256146	mdvnl36_d	899	14	CF256146



```
DT 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
PN
XX 06-JUL-2000.
PD
XX 20-DEC-1999; 99WO-CA001218.
PF
XX 23-DEC-1998; 98US-0113800P.
PR
XX (BIOC-) BIOCHEM PHARMA INC.
PA
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
PI
XX WPI; 2000-452397/39.
XX P-PSDB; AAB12720.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
XX Example 6; Fig 16; 106pp; English.
PS
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
CC bactericidal activity. The nucleic acids, encoding the protein antigens,
CC may be used for the recombinant production of the proteins they encode.
CC The protein antigens may then be used as vaccines for the prevention and
CC treatment of Streptococcal infections in mammals (especially humans)
CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11-2
CC protein antigen
XX
XX Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 53; DB 3; Length 2639;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1035
Db 1113 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1165

RESULT 25
ABK15104
ID ABK15104 standard; DNA; 2639 BP.
XX
XX ABK15104;
AC
XX 08-MAY-2002 (first entry)
DT
XX DNA encoding Streptococcus pneumoniae BVH-11-2.
DE
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
KW streptococcal bacterial infection; gene; ds; BVH-11-2.
XX Streptococcus pneumoniae.
OS
XX Key Location/Qualifiers
FH 114..2630
FT CDS
FT /tag= a
FT /product= "BVH-11-2"
FT /note= "The gene is flanked by sequences from the vector
FT SP64, no information on which is given in the
FT specification"
```

```
XX WO200198334-A2.
PN
XX 27-DEC-2001.
PD
XX 19-JUN-2001; 2001WO-CA000908.
PF
XX 20-JUN-2000; 2000US-0212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
PA
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI
XX WPI; 2002-122272/16.
XX P-PSDB; AAU75934.
DR
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
PT polypeptides, useful as vaccine components for treating or preventing
XX streptococcal infections such as otitis media, meningitis, and
XX bacteraemia.
XX
XX Example 3; Fig 5; 113pp; English.
PS
XX The invention describes an isolated polypeptide (I) with 70-90% identity
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection. A
CC polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence encodes the antigenic
CC Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
XX peptides described in the method of the invention
XX
XX Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 53; DB 6; Length 2639;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1035
Db 1113 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTTGGGTACCAAGTTCAAG 1165

Search completed: September 30, 2004, 23:15:41
Job time : 923 secs
```

```

Db      1003 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1055
|||||
RESULT 22
AAAA7605
ID      AAAA7605 standard; DNA; 2531 BP.
XX
AC      AAAA7605;
XX
XX      20-OCT-2000 (first entry)
XX
DE      Recombinant variant of Sp36 gene (Sp36B) of S. pneumoniae.
XX
KW      Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW      histidine triad residue; Sp36; antibody; otitis media;
KW      nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW      meningitis; lobar pneumonia; ds.
XX
OS      Streptococcus pneumoniae.
XX
FH      Key Location/Qualifiers
FT      CDS 1..2531
FT      /*tag= a
FT      /product= "Sp36B polypeptide"
XX
PN      WO200037105-A2.
XX
XX      29-JUN-2000.
XX
XX      21-DEC-1999; 99WO-US030390.
XX
XX      21-DEC-1998; 98US-0113048P.
XX
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Johnson LS, Koenig S, Adamou JE;
XX
XX      WPI; 2000-452129/39.
XX      P-PSDB; AAB01469.
XX
XX      Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX      such as otitis media, nasopharyngeal and bronchial infections, comprises
XX      Streptococcus pneumoniae proteins.
XX
XX      Disclosure; Page 69-70; 70pp; English.
XX
XX      Although a number of proteins have been suggested as being involved in
XX      the pathogenicity of Streptococcus pneumoniae, there still remains a need
XX      to identify polypeptides having epitopes in common from various strains
XX      of S. pneumoniae in order to utilize such polypeptides in vaccines to
XX      protect against a wide variety of S. pneumoniae. New vaccine compositions
XX      are described which comprise a Streptococcus pneumoniae polypeptide (or
XX      fragments) of 80 - 680 amino acids in length that comprise at least one
XX      histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
XX      directed against these features. The vaccine is useful in protecting
XX      against infection by Streptococcus pneumoniae. The vaccine composition
XX      comprising antibodies to is useful for passive immunization for treating
XX      pneumococcal infections which includes otitis media, nasopharyngeal and
XX      bronchial infections
XX
XX      Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 0 U; 3 Other;

Query Match 2.2%; Score 53; DB 3; Length 2531;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1035
|||||
Db      1000 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1052
|||||
RESULT 23

```

```

AAAA7602
ID      AAAA7602 standard; DNA; 2531 BP.
XX
AC      AAAA7602;
XX
XX      20-OCT-2000 (first entry)
XX
DE      Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.
XX
KW      Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KW      histidine triad residue; Sp36; antibody; otitis media;
KW      nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW      meningitis; lobar pneumonia; ds.
XX
OS      Streptococcus pneumoniae.
XX
FH      Key Location/Qualifiers
FT      CDS 1..2517
FT      /*tag= a
FT      /product= "Sp36D polypeptide"
XX
PN      WO200037105-A2.
XX
XX      29-JUN-2000.
XX
XX      21-DEC-1999; 99WO-US030390.
XX
XX      21-DEC-1998; 98US-0113048P.
XX
XX      (MEDI-) MEDIMMUNE INC.
XX
XX      Johnson LS, Koenig S, Adamou JE;
XX
XX      WPI; 2000-452129/39.
XX      P-PSDB; AAB01466.
XX
XX      Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX      such as otitis media, nasopharyngeal and bronchial infections, comprises
XX      Streptococcus pneumoniae proteins.
XX
XX      Disclosure; Page 57-58; 70pp; English.
XX
XX      Although a number of proteins have been suggested as being involved in
XX      the pathogenicity of Streptococcus pneumoniae, there still remains a need
XX      to identify polypeptides having epitopes in common from various strains
XX      of S. pneumoniae in order to utilize such polypeptides in vaccines to
XX      protect against a wide variety of S. pneumoniae. New vaccine compositions
XX      are described which comprise a Streptococcus pneumoniae polypeptide (or
XX      fragments) of 80 - 680 amino acids in length that comprise at least one
XX      histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody
XX      directed against these features. The vaccine is useful in protecting
XX      against infection by Streptococcus pneumoniae. The vaccine composition
XX      comprising antibodies to is useful for passive immunization for treating
XX      pneumococcal infections which includes otitis media, nasopharyngeal and
XX      bronchial infections
XX
XX      Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 3; Length 2531;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      983 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1035
|||||
Db      1003 GCTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATTTGGTACCAGATTCAAG 1055
|||||
RESULT 24
AAAA7607
ID      AAAA7607 standard; DNA; 2639 BP.
XX
AC      AAAA7607;
XX
XX

```

Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATGGTACCAGATTCAAG 1035  
 |||  
 Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATGGTACCAGATTCAAG 1055

RESULT 20  
 AAA05417  
 ID AAA05417 standard; DNA; 2481 BP.  
 XX  
 AC AAA05417;  
 XX  
 DT 24-MAY-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae nucleotide sequence ID311.  
 XX  
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 KW pneumococcal disease; ds.  
 XX  
 OS Streptococcus pneumoniae.

XX WO200006737-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 XX 27-JUL-1999; 99WO-GB002451.  
 XX  
 PF 27-JUL-1998; 98GB-00016337.  
 PR 19-MAR-1999; 99US-0125164P.  
 XX  
 XX (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 XX Gilbert CFG, Hansbro PW;  
 XX  
 DR WPI; 2000-195300/17.  
 DR P-PSDB; AAY81662.  
 XX  
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of  
 PT pneumococcal diseases and for screening agents capable of antagonizing or  
 PT inhibiting expression of the protein.  
 XX  
 PS Claim 2; Page 99; 108pp; English.

XX AAY81501 to AAY81679 represent specifically claimed protein sequences  
 CC isolated from Streptococcus pneumoniae. AAY05407 to AAY05590 represent  
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
 CC The sequences have antibacterial and antiinflammatory properties. The  
 CC protein sequences, and fragments of them, are useful as immunogens and in  
 CC antigens. The nucleotide sequences can be used in vaccines and in  
 CC diagnostic assays. The proteins and nucleotides can be useful for the  
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
 CC useful for screening an agent capable of antagonising, inhibiting or  
 CC interfering with the function or expression of the proteins in which the  
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
 CC and meningitis. AAY05591 to AAY05614 represent primers used in the  
 CC exemplification of the present invention

XX Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 U; 0 Other;  
 Query Match 2.2%; Score 53; DB 3; Length 2481;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-16;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATGGTACCAGATTCAAG 1035  
 |||  
 Db 1003 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATGGTACCAGATTCAAG 1055

RESULT 21  
 ABX06705  
 ID ABX06705 standard; DNA; 2517 BP.  
 XX  
 AC ABX06705;

XX 27-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain coding region #993.  
 XX  
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 27-MAR-2002; 2002WO-IB002163.  
 PF  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR P-PSDB; ABU01418.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.

XX Claim 6; SEQ ID NO 1985; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56434. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2489  
 CC identified coding region from the genomic sequence. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)

XX Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 7; Length 2517;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-16;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAACACCATGGTACCAGATTCAAG 1035





XX AB084792 to AB084904 represents nucleic acids which encode the  
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.  
CC pneumoniae antigens have antibacterial activity and can be used in  
CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
CC attenuate a Streptococcal infection in an animal. The polynucleotides  
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
CC nucleic acids. AB084905 to AB085130 represent primers used in the cloning  
CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
CC from the present invention  
XX  
SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;

Query Match 2.2%; Score 53; DB 6; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 6.8e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035  
|||  
Db 944 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 996  
|||

RESULT 16  
ID ADC45146 standard; DNA; 2290 BP.  
XX  
AC ADC45146;  
XX

DT 18-DEC-2003 (first entry)

DE S. pneumoniae DNA encoding antigen SP042.

KW Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.

OS Streptococcus pneumoniae.

XX US6573082-B1.

XX 03-JUN-2003.

XX 28-MAR-2000; 2000US-00536784.

XX 31-OCT-1996; 96US-0029960P.

PR 30-OCT-1997; 97US-00961083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;

DR WPI; 2003-764574/72.  
DR P-PSDB; ADC45147.

XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
PT useful for producing vaccines for prevention or attenuation of infection  
PT by Streptococcus pneumoniae.

XX Example 1; SEQ ID NO 65; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a  
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
CC antigens. Also included are making a recombinant vector by inserting the  
CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
CC antigen nucleic acids are useful as probes for use in diagnostic methods  
CC for detecting S. pneumoniae gene expression. The present sequence encodes  
CC an S. pneumoniae antigenic protein.

XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;

Query Match 2.2%; Score 53; DB 9; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 6.8e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 983 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 1035  
|||  
Db 944 GCTCGTATTATCCCTTCGTTATCGTTCAAAACCATGGGTACCAAGATTCAAG 996  
|||

RESULT 17  
AAV52376 standard; DNA; 2359 BP.  
ID AAV52376  
XX  
AC AAV52376;  
XX

DT 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:243.

DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US019588.

XX 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
PI Dougherty BA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.

XX Claim 1; Page 1265-1266; 1409pp; English.

XX The present invention describes a computer readable medium which has the  
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
CC on it, or a representative fragment or a sequence at least 9% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
CC 391, identifying members of the library which contain sequences that  
CC hybridize to the target sequence and isolating the nucleic acid molecules  
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is  
CC homologous to amplification primers derived from the fragment of the S.  
CC pneumoniae genome to prime the amplification and isolating the amplified  
CC sequences. The computer readable medium can be used in a computer-based  
CC system for identifying fragments of the S. pneumoniae genome of  
CC commercial importance, or expression modulating fragments of the S.  
CC pneumoniae genome. Products from the present invention can be used in  
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines  
CC for S. pneumoniae

XX Sequence 2359 BP; 786 A; 451 C; 511 G; 611 T; 0 U; 0 Other;

Query Match 2.2%; Score 53; DB 2; Length 2359;  
Best Local Similarity 100.0%; Pred. No. 6.8e-16;

```
PS Claim 54; Page 52-54; 66pp; English.
XX
CC This DNA sequence encodes a 79 kDa protein (see AAY05753) of
CC Streptococcus pneumoniae serotype 4 that is capable of degrading human
CC complement protein C3 (HCP3). It was identified in the S. pneumoniae
CC serotype 4 genome by alignment to another novel open reading frame (see
CC AAX25393) that codes for a 20 kDa HCP3 protease (AAY05752) of S.
CC pneumoniae serotype 3. This suggested the open reading frame that codes
CC for the 20 kDa protein may be part of a larger open reading frame. Amino
CC acids 1-58 and 90-132 of the 20 kDa protein have substantial sequence
CC identity with amino acids 170-227 and 258-300 of the 79 kDa protein.
CC Proteins and peptides or polypeptides containing these regions, and DNA
CC sequences encoding them (nucleotides 507-681 and 827-999 of the present
CC sequence) are claimed. HCP3 proteases and polypeptides can be used as
CC immune system stimulating compositions (claimed). They can produce an
CC immune response against S. pneumoniae to immunize or treat a mammalian
CC subject against infection or colonization (claimed). They can produce a B
CC cell response, a T cell response, an epithelial cell response, or an
CC endothelial cell response (claimed). The expression of the proteins on
CC the surface of an organ of an animal used in xenotransplantation can be
CC used to inhibit C3-mediated inflammation and rejection
XX
SQ Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 U; 0 Other;
Query Match 2.2%; Score 53; DB 2; Length 2163;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATGGTACGATTCAG 1035
DB 688 GTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATGGTACGATTCAG 740
RESULT 14
AAV27356
ID AAV27356 standard; DNA; 2290 BP.
XX
AC AAV27356;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0042 nucleotide.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 2..2290
FT /tag= a
FT /product= "SP0042"
FT /transl_except= (pos:152..154,aa:Xaa)
FT /transl_except= (pos:1406..1408,aa:Xaa)
FT /transl_except= (pos:1430..1432,aa:Xaa)
FT /note= "no stop codon given; Xaa is unspecified"
XX
BN WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019422.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
XX WPI; 1998-272224/24.
XX
XX F-PSDB; AAW55095.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT
- or their epitope-containing fragments, useful in protective or
therapeutic vaccines, and for diagnosis.
Claim 1; Page 61-62; 118pp; English.
The present sequence encodes a protein from Streptococcus pneumoniae. The
nucleic acid sequence encoding the Streptococcus pneumoniae protein can
be useful in vaccines for inducing protective antibodies against
Streptococcus pneumoniae, for treatment or prevention of infection e.g.
pneumonia, otitis media or meningitis. Probes based on the nucleic acid
are used to detect Streptococcus infection (by usual hybridisation or
amplification methods), also for isolating Streptococcus genes or their
allelic variants. The protein can be used similarly to detect specific
antibodies in standard immunoassays, especially for diagnosing or
monitoring infections. Antibodies which bind the protein are used to
detect corresponding antigens, to purify the protein and for passive
immunisation (optionally coupled to a toxin). Vaccines are administered,
e.g. by injection, orally or through the skin, typically at 0.01-1000
(especially 10-300) mu g/ml per dose
XX
SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;
Query Match 2.2%; Score 53; DB 2; Length 2290;
Best Local Similarity 100.0%; Pred. No. 6.8e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATGGTACGATTCAG 1035
DB 944 GTCGTATTATTCCTCCCTTCGTTATCGTTCAACCATGGTACGATTCAG 996
RESULT 15
ABQ84824
ID ABQ84824 standard; DNA; 2290 BP.
XX
AC ABQ84824;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection; gene; ds.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-00765272.
XX
PR 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2002-479261/51.
XX DR P-PSDB; ABP54589.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
PT infection.
XX
XX Claim 1; Page 28-29; 70pp; English.
```

```
XX WPI; 2002-122272/16.
DR P-PSDB; AAU75933.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
PT polypeptides, useful as vaccine components for treating or preventing
PT streptococcal infections such as otitis media, meningitis, and
PT bacteremia.
XX
XX Disclosure; Fig 4; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90% identity
CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteremia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
CC Staphylococcus aureus) in an individual susceptible to the infection. A
CC polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence encodes the
CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic
CC peptides described in the method of the invention
XX
XX Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;
SQ
Query Match 6.2%; Score 148; DB 6; Length 2647;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CAAGACAACTGGGTACCTTCTGTAAAGCAATCCAGGAACCTACAAATCTAACCAAGCA 758
DB 802 CAAGACAACTGGGTACCTTCTGTAAAGCAATCCAGGAACCTACAAATCTAACCAAGCA 861
QY 759 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAATGACATTGATCTCTTTGAAAC 818
DB 862 ACAACAGCAACACTAACAGTCAAGCAAGTCAAAAGTAATGACATTGATCTCTTTGAAAC 921
QY 819 AGCTCTACAACTGCCCTTTGAGTCAACG 846
DB 922 AGCTCTACAACTGCCCTTTGAGTCAACG 949
RESULT 12
AAA08556
ID AAA08556 standard; DNA; 504 BP.
XX
XX AAA08556;
XX
XX 19-JUL-2000 (first entry)
XX
XX S. pneumoniae 20 kDa human C3-degrading protein coding sequence.
XX
XX Human C3-degrading protein; 20 kDa; immunostimulatory; vaccine;
XX inhibitor; inflammation; organ rejection; xenotransplantation; ss.
XX
XX Streptococcus pneumoniae.
XX
XX WO200017370-A1.
XX
XX 30-MAR-2000.
XX
XX 24-SEP-1999; 99WO-US022362.
XX
XX 24-SEP-1998; 98US-0101736P.
XX
```

```
PR 31-MAR-1999; 99US-00283094.
XX
XX (MINU ) UNIV MINNESOTA.
XX (AMCY ) AMERICAN CYANAMID CO.
XX
XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
XX
XX WPI; 2000-283594/24.
XX P-PSDB; AAY91938.
XX
XX Isolated polypeptide is used to stimulate immune system and immunize or
XX treat a mammalian subject against Streptococcus pneumoniae infection or
XX colonization.
XX
XX Claim 1; Page 53; 63pp; English.
XX
XX The present sequence, isolated from Streptococcus pneumoniae, encodes a
XX human C3-degrading protein (see AAY91938) of about 20 kDa. This sequence
XX may be part of a larger open reading frame (see AAA08557) which encodes
XX an approximately 92 kDa protein also having human C3-degrading activity.
XX The DNA sequences can be used for producing an immune response to
XX Streptococcus pneumoniae in a mammal. Antibodies against the proteins can
XX be used to inhibit S. pneumoniae-mediated C3 degradation. C3-mediated
XX inflammation and rejection in xenotransplantation can be inhibited by
XX expressing the nucleic acid sequences on the surface of an organ of an
XX animal. In particular, the polypeptides are useful for stimulating the
XX immune system and are effective to immunize or treat a mammalian subject
XX against Streptococcus pneumoniae infection or colonization
XX
XX Sequence 504 BP; 153 A; 125 C; 98 G; 128 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 53; DB 3; Length 504;
Best Local Similarity 100.0%; Pred. No. 6.4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 983 GCTCGTATTATTCCTTCGTTATCGTTTCAAAACCATTCGGTACCAGATTCAAG 1035
DB 184 GCTCGTATTATTCCTTCGTTATCGTTTCAAAACCATTCGGTACCAGATTCAAG 236
RESULT 13
AAZ25394
ID AAZ25394 standard; DNA; 2163 BP.
XX
XX AAZ25394;
XX
XX 19-JUL-1999 (first entry)
XX
XX Streptococcus pneumoniae complement C3-degrading protease DNA.
XX
XX Human complement C3-degrading protease; vaccine; infection; meningitis;
XX pneumonia; xenotransplantation; transplant rejection; inflammation; ds.
XX
XX Streptococcus pneumoniae.
XX
XX WO9915675-A1.
XX
XX 01-APR-1999.
XX
XX 24-SEP-1998; 98WO-US020186.
XX
XX 24-SEP-1997; 97US-0059907P.
XX
XX (MINU ) UNIV MINNESOTA.
XX (AMCY ) AMERICAN CYANAMID CO.
XX
XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
XX
XX WPI; 1999-254719/21.
XX P-PSDB; AAY05753.
XX
XX New isolated human complement C3-degrading proteinase.
XX
```

```
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA001218.
XX PR 23-DEC-1998; 98US-0113800P.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX DR P-PSDB; AAB12716.
XX CC Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteremia and/or pneumonia.
XX PS Example 2; Fig 3; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
XX CC bactericidal activity. The nucleic acids, encoding the protein antigens,
XX CC may be used for the recombinant production of the proteins they encode.
XX CC The protein antigens may then be used as vaccines for the prevention and
XX CC treatment of Streptococcal infections in mammals (especially humans)
XX CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
XX CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
XX CC antigen
XX SQ Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 U; 0 Other;
XX CC
Query Match 6.2%; Score 148; DB 3; Length 2523;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATAGTATGATCTCTTTGAAAC 758
Db 758 CAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATAGTATGATCTCTTTGAAAC 817
QY 759 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTTGAAAC 818
Db 818 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTTGAAAC 877
QY 819 AGCTCTACAAACTGCCTTTGAGTCAACG 846
Db 878 AGCTCTACAAACTGCCTTTGAGTCAACG 905
XX CC
RESULT 10
AAA65736
ID AAA65736 standard; DNA; 2647 BP.
XX AC AAA65736;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX OS Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PR 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA001218.
XX PR 23-DEC-1998; 98US-0113800P.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI
```

```
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteremia and/or pneumonia.
XX PS Example 6; Fig 15; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
XX CC bactericidal activity. The nucleic acids, encoding the protein antigens,
XX CC may be used for the recombinant production of the proteins they encode.
XX CC The protein antigens may then be used as vaccines for the prevention and
XX CC treatment of Streptococcal infections in mammals (especially humans)
XX CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
XX CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
XX CC antigen
XX SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;
XX CC
Query Match 6.2%; Score 148; DB 3; Length 2647;
Best Local Similarity 100.0%; Pred. No. 4.9e-64;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATAGTATGATCTCTTTGAAAC 758
Db 802 CAAGAACAAACTGGGTACCTTCTGTAAAGCAATCCAGGAATCAAAATAGTATGATCTCTTTGAAAC 861
QY 759 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTTGAAAC 818
Db 862 ACAACAGCAACTACAGTCAAGCAAGTCAAGTAATGACATTGATGATCTCTTTGAAAC 921
QY 819 AGCTCTACAAACTGCCTTTGAGTCAACG 846
Db 922 AGCTCTACAAACTGCCTTTGAGTCAACG 949
XX CC
RESULT 11
ABK15103
ID ABK15103 standard; DNA; 2647 BP.
XX AC ABK15103;
XX DT 08-MAY-2002 (first entry)
XX DE DNA encoding Streptococcus pneumoniae BVH-11.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
XX KW Streptococcal bacterial infection; gene; ds.
XX OS Streptococcus pneumoniae.
XX FH Key Location/Qualifiers
XX CDS 45..2567
XX FT /tag= a
XX FT /product= "BVH-11"
XX FT /note= "The gene is flanked by sequences from the vector
XX FT SP64, no information on which is given in the
XX FT specification"
XX PN WO200198334-A2.
XX PR 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA000908.
XX PR 20-JUN-2000; 2000US-0212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
```

QY	481	AGGAGCGCTATCTACAGATGATGGTTATATCTTTTAATGCTTCTGTATCATAGAGATAC	540
Db	6470	AGGAGCGCTATCTACAGATGATGGTTATATCTTTTAATGCTTCTGTATCATAGAGATAC	6411
QY	541	TGGTGTAGCTTATATCGTTCCCTCATCGAGATCAATACCATTAATTCCTTAAGAATCAGTT	600
Db	6410	TGGTGTAGCTTATATCGTTCCCTCATCGAGATCAATACCATTAATTCCTTAAGAATCAGTT	6351
QY	601	ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCCCTATCTGGTTCGAGGAATCTGTCAAA	660
Db	6350	ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCCCTATCTGGTTCGAGGAATCTGTCAAA	6291
QY	661	TTCAAGAACCTATCGCCGACAAATAATACGATAACACTTCAAGAACAACTTGGGTGACCTTC	720
Db	6290	TTCAAGAACCTATCGCCGACAAATAATACGATAACACTTCAAGAACAACTTGGGTGACCTTC	6231
QY	721	TGTAAAGCAATCCAGGAACCTAATACTAACAGCAAGCAACAGCAACACTTAACAGTCA	780
Db	6230	TGTAAAGCAATCCAGGAACCTAATACTAACAGCAAGCAACAGCAACACTTAACAGTCA	6171
QY	781	AGCAAGTCAAAAGTAATGACATGTAGTCTCTTTGAAACAGCTCTTCAAACTGCGCTTTGAG	840
Db	6170	AGCAAGTCAAAAGTAATGACATGTAGTCTCTTTGAAACAGCTCTTCAAACTGCGCTTTGAG	6111
QY	841	TCAACGACATGTAGAATCTGATGGCCCTTGCTTTTGATCCAGCAAAATCAAGTCGAAC	900
Db	6110	TCAACGACATGTAGAATCTGATGGCCCTTGCTTTTGATCCAGCAAAATCAAGTCGAAC	6051
QY	901	AGCTTAGAGGTTTGCAGTGCACACGGAGATCAATACCACTTTCATCCCTTACTCTCAAT	960
Db	6050	AGCTTAGAGGTTTGCAGTGCACACGGAGATCAATACCACTTTCATCCCTTACTCTCAAT	5991
QY	961	GTCTGAATTGGAAGAACGAATCGCTCTGATTTATTTCCCTCTTGTTATCGTTTCAAAACCATTC	1020
Db	5990	GTCTGAATTGGAAGAACGAATCGCTCTGATTTATTTCCCTCTTGTTATCGTTTCAAAACCATTC	5931
QY	1021	GGTACACAGATTCGAAGCCAGAAACAAACCAAGTCCACACCGACTCCGAAACCTAGTCCAGG	1080
Db	5930	GGTACACAGATTCGAAGCCAGAAACAAACCAAGTCCACACCGACTCCGAAACCTAGTCCAGG	5871
QY	1081	CCGCAACCTGACACAAATCTTAAATAGACTCAAATCTTTCTTTGGTTAGTCAGCTGGT	1140
Db	5870	CCGCAACCTGACACAAATCTTAAATAGACTCAAATCTTTCTTTGGTTAGTCAGCTGGT	5811
QY	1141	ACGAAAGTTTGGGGAAGGATGTATTTCGAAGAAAAGGGCATCTCTCGTTATCTCTTTGC	1200
Db	5810	ACGAAAGTTTGGGGAAGGATGTATTTCGAAGAAAAGGGCATCTCTCGTTATCTCTTTGC	5751
QY	1201	GAAGAATTTACCATCTGAAACTGTTAAAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG	1260
Db	5750	GAAGAATTTACCATCTGAAACTGTTAAAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG	5691
QY	1261	TGTTTTCACACATTTAACTGCTTAAAAAGAAAATGTTTGCTTCCTGTGACCAAGAAATTTTA	1320
Db	5690	TGTTTTCACACATTTAACTGCTTAAAAAGAAAATGTTTGCTTCCTGTGACCAAGAAATTTTA	5631
QY	1321	TGATAAAGCATATAACTGTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAATTAAGGCTCG	1380
Db	5630	TGATAAAGCATATAACTGTGTTAACTGAGGCTCATAAAGCCCTGTTTGNAATTAAGGCTCG	5571
QY	1381	TAATTCGTATTTCCAGCCCTTAGACAAATTTATTAGAACCGCTTGAATGATGAATCGACTAA	1440
Db	5570	TAATTCGTATTTCCAGCCCTTAGACAAATTTATTAGAACCGCTTGAATGATGAATCGACTAA	5511
QY	1441	TAAAGAAAAATTTGTTAGATGATTTATTGGCAATTCCTTAGACCCAAATTAACCATCCAGAGCG	1500
Db	5510	TAAAGAAAAATTTGTTAGATGATTTATTGGCAATTCCTTAGACCCAAATTAACCATCCAGAGCG	5451
QY	1501	ACTTTGGCAACCAAAATTCCTCAATTTAGTATACTGGAAGACGAAGTTTCGTATTCCTCAATT	1560
Db	5450	ACTTTGGCAACCAAAATTCCTCAATTTAGTATACTGGAAGACGAAGTTTCGTATTCCTCAATT	5391
QY	1561	AGCTGATAAGTATACAACCGTCAGATGGTTTACATTTTTTGATGAACATGATATAATCAAGTGA	1620

Db	5390	AGCTGATAAGTATACAACGTCAGATGGTTACATTTTGGTGAACATGATATATACGTGA	5331
Qy	1621	TGAAGGAGATGCATATCTAAACGCTCATATGGGCCATAGTCACCTGGATTGCGAAAAGATAG	1680
Db	5330	TGAAGGAGATGCATATGTTAAACGCTCATATGGGCCATAGTCACCTGGATTGCGAAAAGATAG	5271
Qy	1681	CTTTCTGTATAAGSAAAAAGTTGCAGCTCAAGCCTATATCTAAAGAAAAAGGTATCCTTACC	1740
Db	5270	CTTTCTGTATAAGSAAAAAGTTGCAGCTCAAGCCTATATCTAAAGAAAAAGGTATCCTTACC	5211
Qy	1741	TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCACAGCTATTTA	1800
Db	5210	TCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCACAGCTATTTA	5151
Qy	1801	CAATCGTGTCAAAGGGGAAAAACGAATTCACACTCGTTTCGACTTCCATATATGTTGAGCA	1860
Db	5150	CAATCGTGTCAAAGGGGAAAAACGAATTCACACTCGTTTCGACTTCCATATATGTTGAGCA	5091
Qy	1861	TACAGTTGAGGTTTAAAAACGGTAAATTTGATTATTCCTCTCACTAAGGATCATTTACCATAATAT	1920
Db	5090	TACAGTTGAGGTTTAAAAACGGTAAATTTGATTATTCCTCTCACTAAGGATCATTTACCATAATAT	5031
Qy	1921	TAAATTTGCTTGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA	1980
Db	5030	TAAATTTGCTTGTTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA	4971
Qy	1981	TTTGTTTGCACGATTAAGTACTACGTPAGAACAACCTGACGAACGCTCCACATTTCTAATGA	2040
Db	4970	TTTGTTTGCACGATTAAGTACTACGTPAGAACAACCTGACGAACGCTCCACATTTCTAATGA	4911
Qy	2041	TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAA	2100
Db	4910	TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAA	4851
Qy	2101	TAAGAACTTCAAAGCGGATGAAGCCGAGTAGAGGAAACACCTGCTGAGCCGAGAAGTCCC	2160
Db	4850	TAAGAACTTCAAAGCGGATGAAGCCGAGTAGAGGAAACACCTGCTGAGCCGAGAAGTCCC	4791
Qy	2161	TCAAGTAGACACTGAAAAAGTAGAGCCCAACTCAAGAAAGCAGAAGTTTGTGTGGCAA	2220
Db	4790	TCAAGTAGACACTGAAAAAGTAGAGCCCAACTCAAGAAAGCAGAAGTTTGTGTGGCAA	4731
Qy	2221	AGTAAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA	2280
Db	4730	AGTAAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA	4671
Qy	2281	TTTGACTCTTCAAATTATGGATAACAAATAGTATCATGTGGCAAGCAGAAAAATTTACTTGC	2340
Db	4670	TTTGACTCTTCAAATTATGGATAACAAATAGTATCATGTGGCAAGCAGAAAAATTTACTTGC	4611
Qy	2341	GTTCGTTAAAAAGGAAGTAATCCTTCATCTGTPAAGTAAAGAAAAATAAAC	2389
Db	4610	GTTCGTTAAAAAGGAAGTAATCCTTCATCTGTPAAGTAAAGAAAAATAAAC	4562

## RESULT 9

AAA65731

ID AAA65731 standard; DNA; 2523 BP.

XX

AC AAA65731;

XX

DT 21-NOV-2000 (first entry)

XX 5

DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.

[illegible]

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KW otitis media; pneumonia; immunisation; bactericidal; ds.  
KW otitis media; pneumonia; immunisation; bactericidal; ds.

XX

OS *Streptococcus pneumoniae*.

XX

PN WO200039299-A2.

QY	1201	GAAAGATTTTACCCTGAAACTGTTAAABATCTTGAAAGCAAGTATTATCAAAACAAGAG	1260
Db	105750	GAAAGATTTTACCCTGAAACTGTTAAABATCTTGAAAGCAAGTATTATCAAAACAAGAG	105691
QY	1261	TGTTTACACACATTTAACTGCTGTAAGAAAGAAATGTTGCTCTCGTGACCAAGATTTTAA	1320
Db	105690	TGTTTACACACATTTAACTGCTGTAAGAAAGAAATGTTGCTCTCGTGACCAAGATTTTAA	105631
QY	1321	TGATAAAGCATATACTGTTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGTCG	1380
Db	105630	TGATAAAGCATATACTGTTTAACTGAGGCTCATAAAGCCCTGTTTGNAAATTAAGGTCG	105571
QY	1381	TAATTCCTGATTTCCAGCCCTTAGACAAATATTAGAACGCTTCAATCATGAATCGACTAA	1440
Db	105570	TAATTCCTGATTTCCAGCCCTTAGACAAATATTAGAACGCTTCAATCATGAATCGACTAA	105511
QY	1441	TAAAGAAAAATTTGGTAGATGATTTATTGGCAATTCCTAGCACCAATTAACCATCCAGAGCG	1500
Db	105510	TAAAGAAAAATTTGGTAGATGATTTATTGGCAATTCCTAGCACCAATTAACCATCCAGAGCG	105451
QY	1501	ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAAGTTCGTATTGCTCAAT	1560
Db	105450	ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAAGTTCGTATTGCTCAAT	105391
QY	1561	AGCTGATAAGTATACAAAGTCAGATGTTTACATTTTGGTGAACATGATATATCAAGTGA	1620
Db	105390	AGCTGATAAGTATACAAAGTCAGATGTTTACATTTTGGTGAACATGATATATCAAGTGA	105331
QY	1621	TGAAGGAGATGCAATATGTAACGGCTCATATGGGCCATAGTCACTGGATTGGAAAAAGATAG	1680
Db	105330	TGAAGGAGATGCAATATGTAACGGCTCATATGGGCCATAGTCACTGGATTGGAAAAAGATAG	105271
QY	1681	CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTAAAGAAAAAGTATCTCTACC	1740
Db	105270	CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTTATCTAAAGAAAAAGTATCTCTACC	105211
QY	1741	TCCATCTCCAGACGCAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCACTATTTA	1800
Db	105210	TCCATCTCCAGACGCAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCACTATTTA	105151
QY	1801	CAATCTGTGAAAGGGGAAAAACGAATTCACCTGCTTCCATATATGTTGGTGAACA	1860
Db	105150	CAATCTGTGAAAGGGGAAAAACGAATTCACCTGCTTCCATATATGTTGGTGAACA	105091
QY	1861	TACAGTTGAGGTTTAAACCGTAAATTTGATTTCTCTCATAGGATCATTAACATTAAT	1920
Db	105090	TACAGTTGAGGTTTAAACCGTAAATTTGATTTCTCTCATAGGATCATTAACATTAAT	105031
QY	1921	TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAA	1980
Db	105030	TAAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAA	104971
QY	1981	TTTGTGTTGGCAATTAAGTACTAOTAGAACCCTGACGAACCTGCCATTTCTAATGA	2040
Db	104970	TTTGTGTTGGCAATTAAGTACTAOTAGAACCCTGACGAACCTGCCATTTCTAATGA	104911
QY	2041	TGATGGGGCAATGCGAGTGAGCTGTTAGCAAGAAAGACCAAGTGAAGATCCAA	2100
Db	104910	TGATGGGGCAATGCGAGTGAGCTGTTAGCAAGAAAGACCAAGTGAAGATCCAA	104851
QY	2101	TAAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	104850	TAAAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	104791
QY	2161	TCAAGTAGAGATGTAAGAAAGTAGAGCCCAATCTAAAGAAAGCAGAGTGTTCCTGCGAA	2220
Db	104790	TCAAGTAGAGATGTAAGAAAGTAGAGCCCAATCTAAAGAAAGCAGAGTGTTCCTGCGAA	104731
QY	2221	AGTAACGGATTTCTAGTCTCAAGCCCAATGCAACAGAAATCTTAGCTGGTTTACGAATAA	2280
Db	104730	AGTAACGGATTTCTAGTCTCAAGCCCAATGCAACAGAAATCTTAGCTGGTTTACGAATAA	104671

QY	2281	TTTGACTCTTCAAAATTTGGATTAACAATAGTATCATGGCAGAGAGAAAAATTTACTTGC	2340
Db	104670	TTTGACTCTTCAAAATTTGGATTAACAATAGTATCATGGCAGAGAGAAAAATTTACTTGC	104611
QY	2341	GTGTTTAAAGGAAGTAATCTCTCATCTGTAAGTAAGGAAAAAATAAAC	2389
Db	104610	GTGTTTAAAGGAAGTAATCTCTCATCTGTAAGTAAGGAAAAAATAAAC	104562

RESULT 8

ABSS56454\_10/c

Continuation (11 of 22) of ABSS56454 from base 1000001 (Streptococcus pneumoniae type 4 st

WP Sequence split into 22 fragments LOCUS ABSS56454 Accession ABSS56454

WP	Fragment Name	Begin	End
WP	ABSS56454_00	1	110000
WP	ABSS56454_01	100001	210000
WP	ABSS56454_02	200001	310000
WP	ABSS56454_03	300001	410000
WP	ABSS56454_04	400001	510000
WP	ABSS56454_05	500001	610000
WP	ABSS56454_06	600001	710000
WP	ABSS56454_07	700001	810000
WP	ABSS56454_08	800001	910000
WP	ABSS56454_09	900001	1010000
WP	ABSS56454_10	1000001	1110000
WP	ABSS56454_11	1100001	1210000
WP	ABSS56454_12	1200001	1310000
WP	ABSS56454_13	1300001	1410000
WP	ABSS56454_14	1400001	1510000
WP	ABSS56454_15	1500001	1610000
WP	ABSS56454_16	1600001	1710000
WP	ABSS56454_17	1700001	1810000
WP	ABSS56454_18	1800001	1910000
WP	ABSS56454_19	1900001	2010000
WP	ABSS56454_20	2000001	2110000
WP	ABSS56454_21	2100001	2162598

Query Match 97.9%; Score 2338; DB 7; Length 110000;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	TTCTTACAGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTTA	60
Db	6950	TTCTTACAGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTTA	6891
QY	61	TATAGATGAAAAAACAAGCGACCCAAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120
Db	6890	TATAGATGAAAAAACAAGCGACCCAAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAA	6831
QY	121	GGGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC	180
Db	6830	GGGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC	6771
QY	181	TTACATGCGCAACCACTATCATTTTAAATGGAAGTTTCCTTATGACGCTATCATCAG	240
Db	6770	TTACATGCGCAACCACTATCATTTTAAATGGAAGTTTCCTTATGACGCTATCATCAG	6711
QY	241	TGAAGAAATTAATCATGAAAGATCCAAAATTAAGCTAAAGATAGGATATTGTTAATGA	300
Db	6710	TGAAGAAATTAATCATGAAAGATCCAAAATTAAGCTAAAGATAGGATATTGTTAATGA	6651
QY	301	GGTCAAGGTTGGATATGTTTCAAGTAGATGAAAAATCTATGTTTACCTTTAAGGATCC	360
Db	6650	GGTCAAGGTTGGATATGTTTCAAGTAGATGAAAAATCTATGTTTACCTTTAAGGATCC	6591
QY	361	TGCCCAACCGGATAACGCTCCGTACAAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG	420
Db	6590	TGCCCAACCGGATAACGCTCCGTACAAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG	6531
QY	421	TCAACATCGTGAAGTGGAACTCCAAGAAAAAGATGGTCTGTTGCTTGGCAAGTTCGCA	480
Db	6530	TCAACATCGTGAAGTGGAACTCCAAGAAAAAGATGGTCTGTTGCTTGGCAAGTTCGCA	6471



Db	4913	TACAGTTGAGGTTAAACCGTAAATTTGATTATTTCTCATAAGGATCAATACCATAATAT	4972
Qy	1921	TAAATTTGCTTGGTTTGAGTATCAACATCAAAAGCTCCAAATGGCTATACCTTTGGAGA	1980
Db	4973	TAAATTTGCTTGGTTTGAGTATCAACATCAAAAGCTCCAAATGGCTATACCTTTGGAGA	5032
Qy	1981	TTTCTTTGCGACGAGTAAAGTAACTAGTAAACACCCCTGACCAACCTGACCAATTTCTAATGA	2040
Db	5033	TTTCTTTGCGACGAGTAAAGTAACTAGTAAACACCCCTGACCAACCTGACCAATTTCTAATGA	5092
Qy	2041	TGGATGGGGCAATGCCAGTGAAGCTGTGTTAGGCAAGAACACACAGTGAAGATCCAAA	2100
Db	5093	TGGATGGGGCAATGCCAGTGAAGCTGTGTTAGGCAAGAACACACAGTGAAGATCCAAA	5152
Qy	2101	TAAAGACTTCAAAGCGGATGAAGACCGTGAAGAACACCTGCTGAGCCAGAGTCCC	2160
Db	5153	TAAAGACTTCAAAGCGGATGAAGACCGTGAAGAACACCTGCTGAGCCAGAGTCCC	5212
Qy	2161	TCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAAGAACAGAAAGTTTGGCTTGGCAA	2220
Db	5213	TCAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAAGAACAGAAAGTTTGGCTTGGCAA	5272
Qy	2221	AGTAACGGATCTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA	2280
Db	5273	AGTAACGGATCTTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA	5332
Qy	2281	TTTGACTCTTCAAATTTATGATTAACATAGTATCATGCGCAGACAGAAATTAATTTGC	2340
Db	5333	TTTGACTCTTCAAATTTATGATTAACATAGTATCATGCGCAGACAGAAATTAATTTGC	5392
Qy	2341	GTTGTTAAAGGAAGTAACTCTTCACTGTAAGTAAGGAAAAATTAAC	2389
Db	5393	GTTGTTAAAGGAAGTAACTCTTCACTGTAAGTAAGGAAAAATTAAC	5441

RESULT 7  
 ABS56454\_09/c  
 Continuation (10 of 22) of ABS56454 from base 900001 (Streptococcus pneumoniae type 4 s)  
 WP Sequence split into 22 fragments  
 WP Fragment Name Begin End  
 WP ABS56454\_00 1 110000  
 WP ABS56454\_01 100001 210000  
 WP ABS56454\_02 200001 310000  
 WP ABS56454\_03 300001 410000  
 WP ABS56454\_04 400001 510000  
 WP ABS56454\_05 500001 610000  
 WP ABS56454\_06 600001 710000  
 WP ABS56454\_07 700001 810000  
 WP ABS56454\_08 800001 910000  
 WP ABS56454\_09 900001 1010000  
 WP ABS56454\_10 1000001 1110000  
 WP ABS56454\_11 1100001 1210000  
 WP ABS56454\_12 1200001 1310000  
 WP ABS56454\_13 1300001 1410000  
 WP ABS56454\_14 1400001 1510000  
 WP ABS56454\_15 1500001 1610000  
 WP ABS56454\_16 1600001 1710000  
 WP ABS56454\_17 1700001 1810000  
 WP ABS56454\_18 1800001 1910000  
 WP ABS56454\_19 1900001 2010000  
 WP ABS56454\_20 2000001 2110000  
 WP ABS56454\_21 2100001 2162598

Query Match 97.9%; Score 2338; DB 7; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGACCGTTAAGGAAAAATATCGTGTTCCTA 60  
 Db 106950 TTCTTACGAGTTGGGACTGTATCAAGCTAGACCGTTAAGGAAAAATATCGTGTTCCTA 106891  
 Qy 61 TATAGATGAAAAACAGCGACGCAAAAAACCGAGAAATTTGACTCTCGTATGAGGTTAGCAA 120

Db	106890	TATAGATGAAAAACAGCGACCAAAAAACCGAGAAATTTGACTCTCGTATGAGGTTAGCAA	106931
Qy	121	GCCTGAAGGATCAATCTGAGCAAAATCGTCAATCAAGATTAACAGCAACCAAGCTATGTCCAC	180
Db	106830	GCCTGAAGGATCAATCTGAGCAAAATCGTCAATCAAGATTAACAGCAACCAAGCTATGTCCAC	106771
Qy	181	TTACATGCGGCACCACTATCAATTTTAAATGGAAGTTCCTTTATGACGCTATCATCAG	240
Db	106770	TTACATGCGGCACCACTATCAATTTTAAATGGAAGTTCCTTTATGACGCTATCATCAG	106711
Qy	241	TGAAGAAATTAATCTATGAAAGATCCAAATTAAGCTAAAGATGAGGATATTTGTAATGA	300
Db	106710	TGAAGAAATTAATCTATGAAAGATCCAAATTAAGCTAAAGATGAGGATATTTGTAATGA	106651
Qy	301	GGTCAAGGTTGGATATGTTATCAAGGATAGTGAAGAAATPACTATGTTTACCTTAAGGATGC	360
Db	106650	GGTCAAGGTTGGATATGTTATCAAGGATAGTGAAGAAATPACTATGTTTACCTTAAGGATGC	106591
Qy	361	TGCCACGCGGATAAACGTTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAAGAGCATAG	420
Db	106590	TGCCACGCGGATAAACGTTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAAGAGCATAG	106531
Qy	421	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTGCTGGCACGTTCCGA	480
Db	106530	TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTGCTGGCACGTTCCGA	106471
Qy	481	AGGACGCTATACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC	540
Db	106470	AGGACGCTATACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGATAC	106411
Qy	541	TGGTGATGCTTATATCGTTTCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	600
Db	106410	TGGTGATGCTTATATCGTTTCTCATGGAGATCATTAACCATTAATTCCTAAGAAATGAGTT	106351
Qy	601	ATCAGTAGGAGTTGGCTGCTGCAAGAGCCTTCTTATCTGGTCAGGAAATCTCTCAA	660
Db	106350	ATCAGTAGGAGTTGGCTGCTGCAAGAGCCTTCTTATCTGGTCAGGAAATCTCTCAA	106291
Qy	661	TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACTTTCAAGAACAACTGGGTACCTTC	720
Db	106290	TTCAAGAACCTTATCGCCGACAAAATAGCGATAACACTTTCAAGAACAACTGGGTACCTTC	106231
Qy	721	TGTAAGCAATCCAGGAATACAAAATTAACAAGCAACAAACAGCAACACCTAACAGTCA	780
Db	106230	TGTAAGCAATCCAGGAATACAAAATTAACAAGCAACAAACAGCAACACCTAACAGTCA	106171
Qy	781	AGCAAGTCAAAAGTAAATGACATTTGATGCTCTTTGAAACAGCTCTCAAACTGCCCTTGAG	840
Db	106170	AGCAAGTCAAAAGTAAATGACATTTGATGCTCTTTGAAACAGCTCTCAAACTGCCCTTGAG	106111
Qy	841	TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC	900
Db	106110	TCAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC	106051
Qy	901	AGCTAGAGTGTTCAGTGCCACACGAGATCAATTAACCATTCATCCCTTACTCTCAAT	960
Db	106050	AGCTAGAGTGTTCAGTGCCACACGAGATCAATTAACCATTCATCCCTTACTCTCAAT	105991
Qy	961	GTCTGAATTTGGAAGAACGATCGCTGTTATTTCCCTTCTGTTATCGTTCAAACTATG	1020
Db	105990	GTCTGAATTTGGAAGAACGATCGCTGTTATTTCCCTTCTGTTATCGTTCAAACTATG	105931
Qy	1021	GGTACACAGATTCAGGCGCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG	1080
Db	105930	GGTACACAGATTCAGGCGCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAG	105871
Qy	1081	CCCGCAACCTGCACCAAAATCTTTAAAATAGACTCAAAATTTCTTTGGTTAGTCAGCTGT	1140
Db	105870	CCCGCAACCTGCACCAAAATCTTTAAAATAGACTCAAAATTTCTTTGGTTAGTCAGCTGT	105811
Qy	1141	ACGAAAAGTTGGGAGGATGTTATTCGAGAAAGGCACTCTCTCGTTATGCTTTTCC	1200
Db	105810	ACGAAAAGTTGGGAGGATGTTATTCGAGAAAGGCACTCTCTCGTTATGCTTTTCC	105751



probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 CC 391, identifying members of the library which contain sequences that  
 CC hybridize to the target sequence and isolating the nucleic acid molecules  
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 CC homologous to amplification primers derived from the fragment of the S.  
 CC pneumoniae genome to prime the amplification and isolating the amplified  
 CC sequences. The computer readable medium can be used in a computer-based  
 CC system for identifying fragments of the S. pneumoniae genome of  
 CC commercial importance, or expression modulating fragments of the S.  
 CC pneumoniae genome. Products from the present invention can be used in  
 CC diagnosis kits and assays, and pharmaceutical compositions and vaccines  
 CC for S. pneumoniae  
 CC  
 XX  
 SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;

Query Match 97.9%; Score 2338; DB 2; Length 8195;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	TTCTTAGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA	60
DB	3053	TTCTTAGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA	3112
QY	61	TATAGATGGAAACAAACGACGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAA	120
DB	3113	TATAGATGGAAACAAACGACGACGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAA	3172
QY	121	CGGTGAAGGAATCAATCTGAGCAATTCGTATCAAGATACAGACCAAGGCTATGTCAC	180
DB	3173	CGGTGAAGGAATCAATCTGAGCAATTCGTATCAAGATACAGACCAAGGCTATGTCAC	3232
QY	181	TTCAATCGGCGACCACTATCATTAATGTAAGTGGTAAAGGTTCCCTTATGACGCTATCATCAG	240
DB	3233	TTCAATCGGCGACCACTATCATTAATGTAAGTGGTAAAGGTTCCCTTATGACGCTATCATCAG	3292
QY	241	TGAAGAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGA	300
DB	3293	TGAAGAATTAATCATGAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGA	3352
QY	301	GGTCAAGGTTGGATATGTTATCAAGGTAGATGMAATATCTATGTTTACCTTAAGGATGC	360
DB	3353	GGTCAAGGTTGGATATGTTATCAAGGTAGATGMAATATCTATGTTTACCTTAAGGATGC	3412
QY	361	TGCCCAACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	420
DB	3413	TGCCCAACGCGGATAACGTCGGTCAAAAGAGGAAATCAATCGACAAACAAAGAGCATAG	3472
QY	421	TCAACATCGTGAAGGTGGAACTCCAGAAACGATGTTGCTGTTGCCCTTGGCAGCTTCGCA	480
DB	3473	TCAACATCGTGAAGGTGGAACTCCAGAAACGATGTTGCTGTTGCCCTTGGCAGCTTCGCA	3532
QY	481	AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC	540
DB	3533	AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC	3592
QY	541	TGTTGATGCTTATATCGTTCCTCATGAGATCATTAACATTACATTCCTCAAGAAATGAGTT	600
DB	3593	TGTTGATGCTTATATCGTTCCTCATGAGATCATTAACATTACATTCCTCAAGAAATGAGTT	3652
QY	601	ATCAGCTAGCGAGTTGGCTGCTCGAAGCCTTCTTATCTGTTGGTCGAGGAAATCTGCAAA	660
DB	3653	ATCAGCTAGCGAGTTGGCTGCTCGAAGCCTTCTTATCTGTTGGTCGAGGAAATCTGCAAA	3712
QY	661	TTCAAGAACCTATCGCGCAAAATAGCGATACACTTTCAAGAACAAACCTGGGTACCTTC	720
DB	3713	TTCAAGAACCTATCGCGCAAAATAGCGATACACTTTCAAGAACAAACCTGGGTACCTTC	3772
QY	721	TGTAAGCAATCCAGGAATCAAAATCTAACAAGCAACCAACAGCAACACTAACAGTCA	780
DB	3773	TGTAAGCAATCCAGGAATCAAAATCTAACAAGCAACCAACAGCAACACTAACAGTCA	3832
QY	781	AGCAAGTCAAGTAATGACATGATGATCTCTTTGAAACAGCTCTACAAACTGCTTTTGAG	840

DB	3833	AGCAAGTCAAGTAATGACATTTGATGCTCTTGAACAGCTCTACAAACTGCTTTTGAG	3892
QY	841	TCAACGACATGTAGAACTCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC	900
DB	3893	TCAACGACATGTAGAACTCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC	3952
QY	901	AGCTAGAGTGTGTCAGTGCACACGAGATCAATACCATTCATCCCTTACTCTCAAA	960
DB	3953	AGCTAGAGTGTGTCAGTGCACACGAGATCAATACCATTCATCCCTTACTCTCAAA	4012
QY	961	GTCTGAATTCGAAGAACGAATCGCTGATATTTATTTCCCTTCTGTTATCGTTCAACCAATG	1020
DB	4013	GTCTGAATTCGAAGAACGAATCGCTGATATTTATTTCCCTTCTGTTATCGTTCAACCAATG	4072
QY	1021	GGTACCAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG	1080
DB	4073	GGTACCAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG	4132
QY	1081	CCGCAACCTGCAACCAATCTTAAATAGACTCAATTTCTTTTGGTTAGTCAGCTGGT	1140
DB	4133	CCGCAACCTGCAACCAATCTTAAATAGACTCAATTTCTTTTGGTTAGTCAGCTGGT	4192
QY	1141	ACGAAAGTTGGGGAAGGATATGTTTTCGAAGAAAGGGGCACTCTCTGTTATGCTTTTC	1200
DB	4193	ACGAAAGTTGGGGAAGGATATGTTTTCGAAGAAAGGGGCACTCTCTGTTATGCTTTTC	4252
QY	1201	GAAAGATTTACCATCTGAAACTCTTTAAAAATCTTGAAGCAAGTTATCAAAACAGAGAG	1260
DB	4253	GAAAGATTTACCATCTGAAACTCTTTAAAAATCTTGAAGCAAGTTATCAAAACAGAGAG	4312
QY	1261	TGTTTCACACACTTTAACTGCTAATAAAGAAATGTTGCTCTCGTGACCAAGAAATTTA	1320
DB	4313	TGTTTCACACACTTTAACTGCTAATAAAGAAATGTTGCTCTCGTGACCAAGAAATTTA	4372
QY	1321	TGATAAAGCATATAATCTGTTAACTGAGGCTATAAAGGCTCTTTGNAATTAAGGCTCG	1380
DB	4373	TGATAAAGCATATAATCTGTTAACTGAGGCTATAAAGGCTCTTTGNAATTAAGGCTCG	4432
QY	1381	TAATTCGTATTTCCAGGCTTTAGACAAATATTTAGAACGCTTGAATGATGAATCGACTAA	1440
DB	4433	TAATTCGTATTTCCAGGCTTTAGACAAATATTTAGAACGCTTGAATGATGAATCGACTAA	4492
QY	1441	TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCATCCAGAGCG	1500
DB	4493	TAAAGAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACCAATTAACCATCCAGAGCG	4552
QY	1501	ACTTGGCAACCAAAATTTCTCAAATTCAGTATATCTGAAGACGAAATTCGTATTCCTCAAT	1560
DB	4553	ACTTGGCAACCAAAATTTCTCAAATTCAGTATATCTGAAGACGAAATTCGTATTCCTCAAT	4612
QY	1561	AGCTGATAAGTATACAACTCAGATGTTTACATTTTGTATGAACATGATATATCACTAGTGA	1620
DB	4613	AGCTGATAAGTATACAACTCAGATGTTTACATTTTGTATGAACATGATATATCACTAGTGA	4672
QY	1621	TGAAGAGATGATATGTAACGCTCATATGGCCATCTGCTGATGATGGAAGATAG	1680
DB	4673	TGAAGAGATGATATGTAACGCTCATATGGCCATCTGCTGATGATGGAAGATAG	4732
QY	1681	CTTTCTGATAAGGAAAAAGTTGCAAGCTCAAGCTATATACTAAAGAAAAAGGTTATCTTACC	1740
DB	4733	CTTTCTGATAAGGAAAAAGTTGCAAGCTCAAGCTATATACTAAAGAAAAAGGTTATCTTACC	4792
QY	1741	TCCATCTCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTCAGAGCTATTTA	1800
DB	4793	TCCATCTCAGACGAGATGTTTAAAGCAATCAACTGGAGATAGTCAGAGCTATTTA	4852
QY	1801	CAATCGTGTGAAGGGAAGAAACGAATTCCTCGTTCCACTTCCATATATGTTGAGCA	1860
DB	4853	CAATCGTGTGAAGGGAAGAAACGAATTCCTCGTTCCACTTCCATATATGTTGAGCA	4912
QY	1861	TACAGTTGAGGTTAAAAACGGTAAATTTGATTTATCTCTATAAGGATCATTAACATAAT	1920

901 AGCTAGAGGCTGTGCGAGTGCACACGAGATCAATACCACCTTCACTCCCTTACTCTCAAAAT 960  
Db |||||  
918 AGCTAGAGGCTGTGCGAGTGCACACGAGATCAATACCACCTTCACTCCCTTACTCTCAAAAT 977  
Qy |||||  
961 GTCTGAATTTGAAGAACGAATCGCTCGTATTATTTCCCTTCTGTTATCGTTCAAAACCAATTG 1020  
Db |||||  
978 GTCTGAATTTGAAGAACGAATCGCTCGTATTATTTCCCTTCTGTTATCGTTCAAAACCAATTG 1037  
Qy |||||  
1021 GGTACAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080  
Db |||||  
1038 GGTACAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1097  
Qy |||||  
1081 CCCGCAACTGCACCAAACTCTTAAATAGACTCAAAATCTCTCTTGGTTAGTCACTGGT 1140  
Db |||||  
1098 CCCGCAACTGCACCAAACTCTTAAATAGACTCAAAATCTCTCTTGGTTAGTCACTGGT 1157  
Qy |||||  
1141 ACGAAAGTTGGGGAAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCCTTGC 1200  
Db |||||  
1158 ACGAAAGTTGGGGAAGGATATGTATTCGAAGAAAGGGCATCTCTCGTTATGTCCTTGC 1217  
Qy |||||  
1201 GAAAGATTTACATCTGAACTGTAAATCTTGAAGCAAGTTTCAAAACAGAGAG 1260  
Db |||||  
1218 GAAAGATTTACATCTGAACTGTAAATCTTGAAGCAAGTTTCAAAACAGAGAG 1277  
Qy |||||  
1261 TGTTCACACACTTTAACTGTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTTA 1320  
Db |||||  
1278 TGTTCACACACTTTAACTGTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTTA 1337  
Qy |||||  
1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTGC 1380  
Db |||||  
1338 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTGC 1397  
Qy |||||  
1381 TAAATCTGATTTCCAGCCTTAGACAAATATTAGAACCTTGAATGATGAATCCACTAA 1440  
Db |||||  
1398 TAAATCTGATTTCCAGCCTTAGACAAATATTAGAACCTTGAATGATGAATCCACTAA 1457  
Qy |||||  
1441 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTAACCCATCCAGAGCG 1500  
Db |||||  
1458 TAAAGAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATTAACCCATCCAGAGCG 1517  
Qy |||||  
1501 ACTTGGCAAAACCAATTTCTCAAATTTGATGATATCTGAAGACGAAGTTCGTATTGCTCAAT 1560  
Db |||||  
1518 ACTTGGCAAAACCAATTTCTCAAATTTGATGATATCTGAAGACGAAGTTCGTATTGCTCAAT 1577  
Qy |||||  
1561 ACTGATAGATATACAGCTGAGATGTTACATTTTGTGATGACATGATATTAATCAGTCA 1620  
Db |||||  
1578 AGCTGATAAGTATACAACTGAGATGTTACATTTTGTGATGAACATGATATATCAGTCA 1637  
Qy |||||  
1621 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1680  
Db |||||  
1638 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAG 1697  
Qy |||||  
1681 CCTTCTGATAAGGAAAAAGTTGACGCTCAAGCCTATATAAGAAAAAGGTTATCCTACC 1740  
Db |||||  
1698 CCTTCTGATAAGGAAAAAGTTGACGCTCAAGCCTATATAAGAAAAAGGTTATCCTACC 1757  
Qy |||||  
1741 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGGAGTAGTGCAGCCTATTTA 1800  
Db |||||  
1758 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGGAGTAGTGCAGCCTATTTA 1817  
Qy |||||  
1801 CAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCCGATTCCTCATATATGTTGAGCA 1860  
Db |||||  
1818 CAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCCGATTCCTCATATATGTTGAGCA 1877  
Qy |||||  
1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATATAT 1920  
Db |||||  
1878 TACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATATAT 1937  
Qy |||||  
1921 TAAATTTGCTGTTGATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1980  
Db |||||  
1938 TAAATTTGCTGTTGATGATCACAATCAAGCTCCAAATGGCTATACCTTTGGAAGA 1997  
Qy |||||  
1981 TTTGTTTGCAGCATTAAGTACTACGTAGAACCCCTGAGAACGTCCTCAATTTCTAATGA 2040  
Qy |||||

1998 TTTGTTTGCAGCATTAAGTACTACGTAGAACCCCTGACGAACGTCACATTTCTAATGA 2057  
Qy |||||  
2041 TGGATGGGCAATGCCAGTGCATGTTTGGCAAGAAAGACCAAGTGAAGATCCAAA 2100  
Db |||||  
2058 TGGATGGGCAATGCCAGTGCATGTTTGGCAAGAAAGACCAAGTGAAGATCCAAA 2117  
Qy |||||  
2101 TAAAGAACTTCAAGCGGATGAAGAGCCAGTGTAGAGAAACACCTGCTGAGCCAGAAAGTCCC 2160  
Db |||||  
2118 TAAAGAACTTCAAGCGGATGAAGAGCCAGTGTAGAGAAACACCTGCTGAGCCAGAAAGTCCC 2177  
Qy |||||  
2161 TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAGCAGAAAGTTTTCCTTTCGCA 2220  
Db |||||  
2178 TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAGCAGAAAGTTTTCCTTTCGCA 2237  
Qy |||||  
2221 AGTAACGGATTTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGTGGTTTACGAATAA 2280  
Db |||||  
2238 AGTAACGGATTTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGTGGTTTACGAATAA 2297  
Qy |||||  
2281 TTTGACTCTTCAAAATTTATGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTC 2340  
Db |||||  
2298 TTTGACTCTTCAAAATTTATGATAACAATAGTATCATGGCAGAGCAGAAAAATTTACTTC 2357  
Qy |||||  
2341 GTTGTAAAAAGGAAGTAAATCTTCTGTAAGTAAAGGAAAAATAAAC 2389  
Db |||||  
2358 GTTGTAAAAAGGAAGTAAATCTTCTGTAAGTAAAGGAAAAATAAAC 2406  
Qy |||||

RESULT 6  
AAV52227  
ID AAV52227 standard; DNA; 8195 BP.  
XX  
AC AAV52227;  
XX  
DT 23-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.  
XX  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9818931-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97WO-US019588.  
XX  
PR 31-OCT-1996; 96US-0029960P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
PI Dougherty BA;  
XX  
WPI; 1998-272225/24.  
XX  
PT Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.  
XX  
PS Claim 1; Page 727-732; 1409pp; English.  
XX  
CC The present invention describes a computer readable medium which has the  
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a

QY 2281 TTGACTCTTCAAAATTATGGATAACAATAGTATCATCGCAGACGAGAAAATTACTTGC 2340  
 Db |||||||  
 QY 2340 TTGACTCTTCAAAATTATGGATAACAATAGTATCATCGCAGACGAGAAAATTACTTGC 2399  
 Db |||||||  
 QY 2341 GTTGTAAAAGGAGTAAATCTCTCATCTGTAAGTAAGGAAAAATAAAC 2389  
 Db |||||||  
 QY 2400 GTTGTAAAAGGAGTAAATCTCTCATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 5  
 ABX06886  
 ID ABX06886 standard; DNA; 2406 BP.  
 XX  
 AC ABX06886;  
 XX  
 DT 27-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain coding region #1174.  
 XX  
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 PN W0200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 XX  
 DR P-PSDB; ABU01598.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 PS Claim 6; SEQ ID NO 2347; 56pp; English.  
 XX

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as ABS56454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying

CC immunodominant proteins. The present sequence is one of the 2489  
 CC identified coding region from the genomic sequence. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2406 BP; 834 A; 461 C; 490 G; 621 T; 0 U; 0 Other;  
 Query Match 97.9%; Score 2338; DB 7; Length 2406;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGAAAAATATCGTTTCTTA 60  
 Db |||||||  
 QY 18 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGAAAAATATCGTTTCTTA 77  
 Db |||||||  
 QY 61 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGGTAGCAA 120  
 Db |||||||  
 QY 78 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGGTAGCAA 137  
 Db |||||||  
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180  
 Db |||||||  
 QY 138 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 197  
 Db |||||||  
 QY 181 TTCACATGGCGACCACTATCATTTTACAATGTAAGTTCCTTATGACGCTATCATCAG 240  
 Db |||||||  
 QY 198 TTCACATGGCGACCACTATCATTTTACAATGTAAGTTCCTTATGACGCTATCATCAG 257  
 Db |||||||  
 QY 241 TGAAGAAATTACTCATGAAAGATCCAAACTATAAGCTAAGAGATGAGGATATCTTAATGA 300  
 Db |||||||  
 QY 258 TGAAGAAATTACTCATGAAAGATCCAAACTATAAGCTAAGAGATGAGGATATCTTAATGA 317  
 Db |||||||  
 QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC 360  
 Db |||||||  
 QY 318 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC 377  
 Db |||||||  
 QY 361 TGCCCGACGGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGATAG 420  
 Db |||||||  
 QY 378 TGCCCGACGGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGATAG 437  
 Db |||||||  
 QY 421 TCAACATCTGGAAGTGGAACTCCAAAGAAACGATGCTGCTGTTGCCCTTGSCAGCTTCGCA 480  
 Db |||||||  
 QY 438 TCAACATCTGGAAGTGGAACTCCAAAGAAACGATGCTGCTGTTGCCCTTGSCAGCTTCGCA 497  
 Db |||||||  
 QY 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGTATATCATAGAGGATAC 540  
 Db |||||||  
 QY 498 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGTATATCATAGAGGATAC 557  
 Db |||||||  
 QY 541 TGTGTATGCTTATATCGTTCCCTCATGAGATCATTAACCATTTACATTCCTTAAGGATGAGTT 600  
 Db |||||||  
 QY 558 TGTGTATGCTTATATCGTTCCCTCATGAGATCATTAACCATTTACATTCCTTAAGGATGAGTT 617  
 Db |||||||  
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAAGGCCITTCCTATCTGCTGCGAGGAAATCTGTCAA 660  
 Db |||||||  
 QY 618 ATCAGCTAGCGAGTTGGCTGCTGCAAGGCCITTCCTATCTGCTGCGAGGAAATCTGTCAA 677  
 Db |||||||  
 QY 661 TTCAAGAACCTATCGCCGCAAAAATAGCGATACACTTCAAGAAACAACTGGGTACCTTC 720  
 Db |||||||  
 QY 678 TTCAAGAACCTATCGCCGCAAAAATAGCGATACACTTCAAGAAACAACTGGGTACCTTC 737  
 Db |||||||  
 QY 721 TGTAAACAATCCAGGAACCTACAAATTAACCAAGCAACCAACCAACTAACAGTCA 780  
 Db |||||||  
 QY 738 TGTAAACAATCCAGGAACCTACAAATTAACCAAGCAACCAACCAACTAACAGTCA 797  
 Db |||||||  
 QY 781 AGCAAGTCAAAGTAAATGACATTTAGTAGTCTCTTTGAAACAGCTCTACAACTGCTTTGAG 840  
 Db |||||||  
 QY 798 AGCAAGTCAAAGTAAATGACATTTAGTAGTCTCTTTGAAACAGCTCTACAACTGCTTTGAG 857  
 Db |||||||  
 QY 841 TCAACGACATGTAGAAATCTGATGGGCTTCTTTTGTATCCAGCAAAAATCACAAGTCAAC 900  
 Db |||||||  
 QY 858 TCAACGACATGTAGAAATCTGATGGGCTTCTTTTGTATCCAGCAAAAATCACAAGTCAAC 917  
 Db |||||||

Db 120 TATAGTGGAAAAAAGCGACGCAAAAAAAGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 179  
Qy 121 GCGTGAAGGAATCAATGCTGACCAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180  
Db 180 GCGTGAAGGAATCAATGCTGACCAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 239  
Qy 181 TTCATCGCGACCACTATCATATTATCAATGGTAAGGTTCTTATGACGCTATCATCAG 240  
Db 240 TTCATCGCGACCACTATCATATTATCAATGGTAAGGTTCTTATGACGCTATCATCAG 299  
Qy 241 TGAAGATTTACTCATGAAGATCCAAACTATAAGCTAAAGATAGGATATGTTAAATGA 300  
Db 300 TGAAGATTTACTCATGAAGATCCAAACTATAAGCTAAAGATAGGATATGTTAAATGA 359  
Qy 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTTAAAGGATGC 360  
Db 360 GGTCAAGGTTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTTAAAGGATGC 419  
Qy 361 TGCCACACGCGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420  
Db 420 TGCCACACGCGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 479  
Qy 421 TCAACATCGTAAGGTGAACTCCAAAGAAACGATGGTCTGTTGGCTTGGCAAGTTGCA 480  
Db 480 TCAACATCGTAAGGTGAACTCCAAAGAAACGATGGTCTGTTGGCTTGGCAAGTTGCA 539  
Qy 481 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTCGTATCATAGAGGATAC 540  
Db 540 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTCGTATCATAGAGGATAC 599  
Qy 541 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACCATTAACATTCCTTAAGAAATGAGTT 600  
Db 600 TGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACCATTAACATTCCTTAAGAAATGAGTT 659  
Qy 601 ATCAGCTAGCGATGGTGTGCGAAGCCTTCCTATCTGTGCGAGGAAATCTGTCAA 660  
Db 660 ATCAGCTAGCGATGGTGTGCGAAGCCTTCCTATCTGTGCGAGGAAATCTGTCAA 719  
Qy 661 TTCAGAACCTATCCGCGACAAAATAGCGATAACACTTCAAGAACAACTCGGTACCTTC 720  
Db 720 TTCAGAACCTATCCGCGACAAAATAGCGATAACACTTCAAGAACAACTCGGTACCTTC 779  
Qy 721 TTGAAGCAATCCAGGAATACAAATACATAACAAAGCAACACAGCAACACTAAACAGTCA 780  
Db 780 TTGAAGCAATCCAGGAATACAAATACATAACAAAGCAACACAGCAACACTAAACAGTCA 839  
Qy 781 AGCAAGTCAAAGTAATGATGATGCTCTTTGAAACAGCTCTACAACTGCTGCTTGG 840  
Db 840 AGCAAGTCAAAGTAATGATGATGCTCTTTGAAACAGCTCTACAACTGCTGCTTGG 899  
Qy 841 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACAAATCACAGTCCGAC 900  
Db 900 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACAAATCACAGTCCGAC 959  
Qy 901 AGCTAGAGGTGTGCGAGTCCACACGGAGATCATTAACCACTTCACTCCCTTACTCTCAAT 960  
Db 960 AGCTAGAGGTGTGCGAGTCCACACGGAGATCATTAACCACTTCACTCCCTTACTCTCAAT 1019  
Qy 961 GTCTGAATTTGGAAGAACGATCGCTGTTATTTATTCCTTTCCTTTCGTTTCAAAACCATTTG 1020  
Db 1020 GTCTGAATTTGGAAGAACGATCGCTGTTATTTATTCCTTTCCTTTCGTTTCAAAACCATTTG 1079  
Qy 1021 GTTACCAAGTTCAAGCCAGAACCAAGTCCACAAACGCTCCGGAACCTAGTCCAGG 1080  
Db 1080 GTTACCAAGTTCAAGCCAGAACCAAGTCCACAAACGCTCCGGAACCTAGTCCAGG 1139  
Qy 1081 CCCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGGT 1140  
Db 1140 CCCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGGT 1199  
Qy 1141 ACGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCACTCTCGTTATGCTCTTTCG 1200  
Db 1200 ACGAAAGTTGGGAGGATATGTTATCGAAGAAAGGCACTCTCTCGTTATGCTCTTTCG 1259

Qy 1201 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1260  
Db 1260 GAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1319  
Qy 1261 TGTTTCAACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTTA 1320  
Db 1320 TGTTTCAACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTTA 1379  
Qy 1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTGTTGNAATTAAGGGTGG 1380  
Db 1380 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTGTTGNAATTAAGGGTGG 1439  
Qy 1381 TAAATTCATGTTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440  
Db 1440 TAAATTCATGTTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1499  
Qy 1441 TAAAGAAATTTGGTAGATGATTTTATGCGCATTCCTAGCACCAATTCCTCATCCAGAGG 1500  
Db 1500 TAAAGAAATTTGGTAGATGATTTTATGCGCATTCCTAGCACCAATTCCTCATCCAGAGG 1559  
Qy 1501 ACTTGGCAAAACCAATCTCRAAATGAGTATATCTGAAGACGAAAGTTGCTATTGCTCAAT 1560  
Db 1560 ACTTGGCAAAACCAATCTCRAAATGAGTATATCTGAAGACGAAAGTTGCTATTGCTCAAT 1619  
Qy 1561 AGCTGATTAAGTATACAAAGCTCAGATGGTTTACATTTTTTGATGAACATGATATATCAGTGA 1620  
Db 1620 AGCTGATTAAGTATACAAAGCTCAGATGGTTTACATTTTTTGATGAACATGATATATCAGTGA 1679  
Qy 1621 TGAAGAGATGATATGTAACCCCTCATATGCGCCATAGTCTAGTCTGATTTGGAAGAGATAG 1680  
Db 1680 TGAAGAGATGATATGTAACCCCTCATATGCGCCATAGTCTAGTCTGATTTGGAAGAGATAG 1739  
Qy 1681 CCTTCTGATAGGAAAGTTGCGAGCTCAAGCCTATATCTAAGAAAAAGGTATCCTACC 1740  
Db 1740 CCTTCTGATAGGAAAGTTGCGAGCTCAAGCCTATATCTAAGAAAAAGGTATCCTACC 1799  
Qy 1741 TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGTATTTTA 1800  
Db 1800 TCCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGTATTTTA 1859  
Qy 1801 CAATCGTGTGAAGGGGAAAAACGAATTCCTCGTTGAGCTTCATATATGTTGGTGGAGCA 1860  
Db 1860 CAATCGTGTGAAGGGGAAAAACGAATTCCTCGTTGAGCTTCATATATGTTGGTGGAGCA 1919  
Qy 1861 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTTATCTCATAGGATCATTAACATATAT 1920  
Db 1920 TACAGTTGAGGTTTAAAGCGGTAAATTTGATTTATCTCATAGGATCATTAACATATAT 1979  
Qy 1921 TAAATTTGCTTGGTTTGGATGATCACAATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980  
Db 1980 TAAATTTGCTTGGTTTGGATGATCACAATACAAAGCTCCAAATGGCTATACCTTGGAGA 2039  
Qy 1981 TTTGTTTGGAGCATTAAGTACTAGTAGAAACCCCTGACGAAGCTCCCATTTCTAATGA 2040  
Db 2040 TTTGTTTGGAGCATTAAGTACTAGTAGAAACCCCTGACGAAGCTCCCATTTCTAATGA 2099  
Qy 2041 TGGATGGGCAATGCCAGTGAGCATGTTTAGCGAAGAAAGACCAAGTGAAGATCCAAA 2100  
Db 2100 TGGATGGGCAATGCCAGTGAGCATGTTTAGCGAAGAAAGACCAAGTGAAGATCCAAA 2159  
Qy 2101 TAAAGAACTTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
Db 2160 TAAAGAACTTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2219  
Qy 2161 TCAAGTAGAGATGAAAAAGTAGAGCCCACTCAAGAGACGAGATTTGCTTCGGA 2220  
Db 2220 TCAAGTAGAGATGAAAAAGTAGAGCCCACTCAAGAGACGAGATTTGCTTCGGA 2279  
Qy 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAAACTTAGCTGGTTTACGAATTA 2280  
Db 2280 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAAACTTAGCTGGTTTACGAATTA 2339

QY 1261 TGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320  
 Db 1261 TGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320  
 QY 1321 TGAATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTGNAATAAAGGGTCG 1380  
 Db 1321 TGAATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTTGNAATAAAGGGTCG 1380  
 QY 1381 TAATTCGTATTCGAAGCCTTAGACAATATATAGAACGCTTGAATGATGAATCGACTAA 1440  
 Db 1381 TAATTCGTATTCGAAGCCTTAGACAATATATAGAACGCTTGAATGATGAATCGACTAA 1440  
 QY 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCAATCTAGCACCAATACCATCCAGAGCG 1500  
 Db 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCAATCTAGCACCAATACCATCCAGAGCG 1500  
 QY 1501 ACTTGGCAAAACCAATTTCTCAATTTAGTATATCTGAAGACGAAGTTCGTATTCCTCAAT 1560  
 Db 1501 ACTTGGCAAAACCAATTTCTCAATTTAGTATATCTGAAGACGAAGTTCGTATTCCTCAAT 1560  
 QY 1561 AGCTGATAGTATACACAGCTCAGATGCTTACATTTTGTAGAACATGATATATCACTGA 1620  
 Db 1561 AGCTGATAGTATACACAGCTCAGATGCTTACATTTTGTAGAACATGATATATCACTGA 1620  
 QY 1621 TGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTGGATTCGAAAAGATAG 1680  
 Db 1621 TGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTGGATTCGAAAAGATAG 1680  
 QY 1681 CCTTCTGTAGTAAGAAAAAGTTGAGCTCAGCTAAGGCTATATCTAAAGAAAAAGGTATCTTACC 1740  
 Db 1681 CCTTCTGTAGTAAGAAAAAGTTGAGCTCAGCTAAGGCTATATCTAAAGAAAAAGGTATCTTACC 1740  
 QY 1741 TCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTA 1800  
 Db 1741 TCCATCTCCAGACGCGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTA 1800  
 QY 1801 CAATCGTGTGAAAGGGAAAAAGAAATCCACTCGTTCGACTCCATATATGTTGAGCA 1860  
 Db 1801 CAATCGTGTGAAAGGGAAAAAGAAATCCACTCGTTCGACTCCATATATGTTGAGCA 1860  
 QY 1861 TACAGTTGAGTTAAACCGGTATTTGATATTTCTCATAGATCATTTACCATATAT 1920  
 Db 1861 TACAGTTGAGTTAAACCGGTATTTGATATTTCTCATAGATCATTTACCATATAT 1920  
 QY 1921 TAAATTTGCTTTGTTGATCATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980  
 Db 1921 TAAATTTGCTTTGTTGATCATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980  
 QY 1981 TTTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAACGTCACATTTCTAATGA 2040  
 Db 1981 TTTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAACGTCACATTTCTAATGA 2040  
 QY 2041 TGAATGGGCAATGCCAGTGAGTATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100  
 Db 2041 TGAATGGGCAATGCCAGTGAGTATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100  
 QY 2101 TAAGAACTTCAAAGCGATCAAGACGAGTACGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
 Db 2101 TAAGAACTTCAAAGCGATCAAGACGAGTACGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
 QY 2161 TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAGCAGAGGTTTGTCTTGGCAA 2220  
 Db 2161 TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAGCAGAGGTTTGTCTTGGCAA 2220  
 QY 2221 AGTAACGGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGTTTACGAATAA 2280  
 Db 2221 AGTAACGGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGTTTACGAATAA 2280  
 QY 2281 TTTGACTCTTCAAATATATGATACAAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2340  
 Db 2281 TTTGACTCTTCAAATATATGATACAAATAGTATCATGGCAGAGCAGAAAAATTTACTTGC 2340  
 QY 2341 GTTGTAAAAGGAAGTAATCCTTCATCTCTGTAAGTAAGGAAAAAATAAAC 2389

Db 2341 GTTGTAAAAGGAAGTAATCCTTCATCTCTGTAAGTAAGGAAAAAATAAAC 2389  
 RESULT 4  
 ID AAA47604 standard; DNA; 2451 BP.  
 AC AAA47604;  
 XX 20-OCT-2000 (first entry)  
 XX Recombinant variant of Sp36 gene (Sp36A) of S. pneumoniae.  
 XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;  
 KW histidine triad residue; Sp36; antibody; otitis media;  
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;  
 KW meningitis; lobar pneumonia; ds.  
 XX Streptococcus pneumoniae.  
 XX Key Location/Qualifiers  
 FH 1. .2451  
 CDS /tag= a  
 FT /product= "Sp36A polypeptide"  
 XX WO200037105-A2.  
 XX 29-JUN-2000.  
 XX 21-DEC-1999; 99WO-US030390.  
 XX 21-DEC-1998; 98US-0113048P.  
 XX (MEDI-) MEDIMUNE INC.  
 XX Johnson LS, Koenig S, Adamou JE;  
 WPI; 2000-452129/39.  
 DR P-PSDB; AAB01468.  
 XX Vaccine useful for prophylaxis and treatment of pneumococcal infections  
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises  
 PT Streptococcus pneumoniae proteins.  
 XX Disclosure; Page 64-65; 70pp; English.  
 CC Although a number of proteins have been suggested as being involved in  
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need  
 CC to identify polypeptides having epitopes in common from various strains  
 CC of S. pneumoniae in order to utilize such polypeptides in vaccines to  
 CC protect against a wide variety of S. pneumoniae. New vaccine compositions  
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or  
 CC fragments) of 80 - 680 amino acids in length that comprise at least one  
 CC histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody  
 CC directed against these features. The vaccine is useful in protecting  
 CC against infection by Streptococcus pneumoniae. The vaccine composition  
 CC comprising antibodies to is useful for passive immunization for treating  
 CC Pneumococcal infections which includes otitis media, nasopharyngeal and  
 CC bronchial infections  
 SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 0 U; 1 Other;  
 Query Match 100.0%; Score 2389; DB 3; Length 2451;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTCTTACGAGTTGGGACTGCTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA 60  
 Db 60 TTCTTACGAGTTGGGACTGCTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA 119  
 QY 61 TATAGATGAAAAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTCATGAGTTAGCAA 120

QY 2341 GTTGTAAAGGAAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389  
 DB 2341 GTTGTAAAGGAAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389

RESULT 3  
 ID ADC45136 standard; DNA; 2389 BP.  
 AC ADC45136;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE S. pneumoniae DNA encoding antigen SP036.  
 XX  
 KW Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN US6573082-B1.  
 XX  
 PD 03-JUN-2003.  
 XX  
 PF 28-MAR-2000; 2000US-00536784.  
 XX  
 PR 31-OCT-1996; 96US-0029960P.  
 PR 30-OCT-1997; 97US-00961083.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 PI  
 DR WPI; 2003-764574/72.  
 DR P-PSDB; ADC45137.  
 XX  
 XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
 PT useful for producing vaccines for prevention or attenuation of infection  
 PT by Streptococcus pneumoniae.  
 XX  
 PS Example 1; SEQ ID NO 55; 58pp; English.  
 CC  
 CC The invention relates to an isolated polynucleotide consisting of a  
 CC Streptococcus pneumoniae nucleic acid (appearing as ABC45122 and encoding  
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
 CC antigens. Also included are making a recombinant vector by inserting the  
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a  
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
 CC antigen nucleic acids are useful as probes for use in diagnostic methods  
 CC for detecting S. pneumoniae gene expression. The present sequence encodes  
 CC an S. pneumoniae antigenic protein.  
 XX  
 SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Query Match 100.0%; Score 2389; DB 9; Length 2389;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 60  
 DB 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 60  
 QY 61 TATAGATGAAACAAAGCGACGCCAAAAACCGAGAAATTTCATCTCGATGAGGTTAGCAA 120  
 DB 61 TATAGATGAAACAAAGCGACGCCAAAAACCGAGAAATTTCATCTCCTGATGAGGTTAGCAA 120  
 QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAGATAACAGACCAGGCTATGTCAC 180  
 DB 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAGATAACAGACCAGGCTATGTCAC 180

QY 181 TTCATATGCGACCACTATCAATTAATCAATGGTAAGGTTCTTATGACGCTATCATCAG 240  
 DB 181 TTCATATGCGACCACTATCAATTAATCAATGGTAAGGTTCTTATGACGCTATCATCAG 240  
 QY 241 TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATAGGATATTTGTTAAAGA 300  
 DB 241 TGAAGAATTACTCATGAAAGATCCAACTATAAGCTAAAGATAGGATATTTGTTAAAGA 300  
 QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGATGC 360  
 DB 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGATGC 360  
 QY 361 TGCCCAACGGGATTAACGTCGTTACAAAAGAGAGAAATCAATCGACAAAAACAAGAGCATAG 420  
 DB 361 TGCCCAACGGGATTAACGTCGTTACAAAAGAGAGAAATCAATCGACAAAAACAAGAGCATAG 420  
 QY 421 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACGATGGTCTGCTGGCCTTGGCACTGCGCA 480  
 DB 421 TCAACATCGTGAAGGTGGAACCTCCAAAGAAACGATGGTCTGCTGGCCTTGGCACTGCGCA 480  
 QY 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540  
 DB 481 AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540  
 QY 541 TGGTGATGCTTATATCGTTCCCTCATGGAGATCATTAACCATTAACATTCCTAAGAATGAGTT 600  
 DB 541 TGGTGATGCTTATATCGTTCCCTCATGGAGATCATTAACCATTAACATTCCTAAGAATGAGTT 600  
 QY 601 ATCAGTAGGAGTTGGCTGCTGAGAGAGCTTCTATCTGCTGAGGAAATCTGTCAAA 660  
 DB 601 ATCAGTAGGAGTTGGCTGCTGAGAGAGCTTCTATCTGCTGAGGAAATCTGTCAAA 660  
 QY 661 TTCAGAAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAAACAACTCGGTCTCCTTC 720  
 DB 661 TTCAGAAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAAACAACTCGGTCTCCTTC 720  
 QY 721 TGTAAAGCAATCCAGGAACCTACAAATATAACAACAAGCAACAACAGCAACACTTAACAGTCA 780  
 DB 721 TGTAAAGCAATCCAGGAACCTACAAATATAACAACAAGCAACAACAGCAACACTTAACAGTCA 780  
 QY 781 AGCAAGTCAAGTAATGATGATGATCTCTTTGAAAACAGCTCTACAACTCCCTTTGAG 840  
 DB 781 AGCAAGTCAAGTAATGATGATGATCTCTTTGAAAACAGCTCTACAACTCCCTTTGAG 840  
 QY 841 TCAACGACATGATGATGATCTCTTTGATCCAGCAACAATCACAAGTCCGAAC 900  
 DB 841 TCAACGACATGATGATGATCTCTTTGATCCAGCAACAATCACAAGTCCGAAC 900  
 QY 901 AGCTAGAGGTGTGCGAGTGCCACACGAGATCATTAACCACTTCTCCCTTACTCTCAAAAT 960  
 DB 901 AGCTAGAGGTGTGCGAGTGCCACACGAGATCATTAACCACTTCTCCCTTACTCTCAAAAT 960  
 QY 961 GTCTGAAATTGGAAGAACGAAATCGCTGATATTTCCCTTCTGTTTATCGTTTCAACCATG 1020  
 DB 961 GTCTGAAATTGGAAGAACGAAATCGCTGATATTTCCCTTCTGTTTATCGTTTCAACCATG 1020  
 QY 1021 GGTACAGATTCAAGCCGAGAACCAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080  
 DB 1021 GGTACAGATTCAAGCCGAGAACCAACAGTCCCAACCGACTCCGGAACCTAGTCCAGG 1080  
 QY 1081 CCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTCTTTGGTGTAGTCAGTGT 1140  
 DB 1081 CCGCAACCTGCAACCAATCTTTAAATAGACTCAAAATCTCTTTGGTGTAGTCAGTGT 1140  
 QY 1141 AGCAAAAGTTGGGAGAGATATGTTTGAAGAAAGGCACTCTCTGTTATGTTCTTTC 1200  
 DB 1141 AGCAAAAGTTGGGAGAGATATGTTTGAAGAAAGGCACTCTCTGTTATGTTCTTTC 1200  
 QY 1201 GAAAGATTACCATCTGAAACTGTTTAAATACTCTGAAGCAAGTATCAAAAACAAGAGAG 1260  
 DB 1201 GAAAGATTACCATCTGAAACTGTTTAAATACTCTGAAGCAAGTATCAAAAACAAGAGAG 1260



Db 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180  
QY 181 TTCACATGGCCAGCACATCAATTAATACAACTGGTAAGGTTCCCTATGACGCTATCATCAG 240  
Db 181 TTCACATGGCCAGCACATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240  
QY 241 TGAAGAAATTAATCAATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTAATGA 300  
Db 241 TGAAGAAATTAATCAATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTAATGA 300  
QY 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATTAATTAATTAATTAATTAATTA 360  
Db 301 GGTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATTAATTAATTAATTAATTAATTA 360  
QY 361 TGCCCAAGGATACAGTCCCTCAAGAGAGAAATCAATCGACAAAGAGCATAG 420  
Db 361 TGCCCAAGGATACAGTCCCTCAAGAGAGAAATCAATCGACAAAGAGCATAG 420  
QY 421 TCAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTCTGCTTGGCCAGCTTCGCA 480  
Db 421 TCAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTCTGCTTGGCCAGCTTCGCA 480  
QY 481 AGGACGCTATCACTACAGATGATGTTATCTTTAATGCTTCTCATATCATAGAGGATAC 540  
Db 481 AGGACGCTATCACTACAGATGATGTTATCTTTAATGCTTCTCATATCATAGAGGATAC 540  
QY 541 TGGTGATGCTTATATCGTTCCTCATGAGATCATACCATTAATGCTTCTTAAGATGATTT 600  
Db 541 TGGTGATGCTTATATCGTTCCTCATGAGATCATACCATTAATGCTTCTTAAGATGATTT 600  
QY 601 ATCAGCTAGCGAGTGGCTGTCGAGAGCCCTTCTATCTGGTCGAGAAATCTGTCAA 660  
Db 601 ATCAGCTAGCGAGTGGCTGTCGAGAGCCCTTCTATCTGGTCGAGAAATCTGTCAA 660  
QY 661 TTCAAGAACCTATCGCCGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720  
Db 661 TTCAAGAACCTATCGCCGACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720  
QY 721 TGTAAAGCAATCCAGGAATCAATATCTAACACAGCAACAAACAGCAACCTAACAGTCA 780  
Db 721 TGTAAAGCAATCCAGGAATCAATATCTAACACAGCAACAAACAGCAACCTAACAGTCA 780  
QY 781 AGCAAGTCAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAACTGGCTTTGAG 840  
Db 781 AGCAAGTCAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAACTGGCTTTGAG 840  
QY 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCAAAATCAAGTCAAGC 900  
Db 841 TCAACGACATGTAGAATCTGATGGCTTGTCTTGTATCCAGCAAAATCAAGTCAAGC 900  
QY 901 AGCTAGAGGTGTTGCAAGTGCACAGGAGATCATTAACCACTTCTTCTTCTTCTTCTTCTT 960  
Db 901 AGCTAGAGGTGTTGCAAGTGCACAGGAGATCATTAACCACTTCTTCTTCTTCTTCTTCTT 960  
QY 961 GTCTGAATTTGGAAGAACGAATCGCTGTAATTAATTTCCCTTCTGTTATCGTTCAACCATTC 1020  
Db 961 GTCTGAATTTGGAAGAACGAATCGCTGTAATTAATTTCCCTTCTGTTATCGTTCAACCATTC 1020  
QY 1021 GGTACAGATTTCAAGGCCAGAAACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGG 1080  
Db 1021 GGTACAGATTTCAAGGCCAGAAACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGG 1080  
QY 1081 CCGCAACCTGCAACCAATCTTAAATPAGACTCAAATTTCTTCTTCTTCTTCTTCTTCTTCTT 1140  
Db 1081 CCGCAACCTGCAACCAATCTTAAATPAGACTCAAATTTCTTCTTCTTCTTCTTCTTCTTCTT 1140  
QY 1141 ACAGAAAGTTGGGAAGGATATGTTATTCGAAGAAAGGGGATCTCTGTTATGCTTTTTC 1200  
Db 1141 ACAGAAAGTTGGGAAGGATATGTTATTCGAAGAAAGGGGATCTCTGTTATGCTTTTTC 1200  
QY 1201 GAAAGATTTACCATCTGAACTGTTAAATCTTTGAAAGCAAGTTATCAAAAACAGAGAG 1260

Db 1201 GAAAGATTTACCATCTGAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAGAGAG 1260  
QY 1261 TGTTTCAACACATTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCACCAAGAAATTTTA 1320  
Db 1261 TGTTTCAACACATTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCACCAAGAAATTTTA 1320  
QY 1321 TGTAARAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTAATAAGGGTGC 1380  
Db 1321 TGTAARAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTAATAAGGGTGC 1380  
QY 1381 TAAATTCGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATGATCAATCACTAA 1440  
Db 1381 TAAATTCGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATGATCAATCACTAA 1440  
QY 1441 TAAAGAAAAATTTGTTAGATGATTTATTGGCATTCTTAGCACCATTACCCATCCAGAGCG 1500  
Db 1441 TAAAGAAAAATTTGTTAGATGATTTATTGGCATTCTTAGCACCATTACCCATCCAGAGCG 1500  
QY 1501 ACTTGGCAAAACCAAAATCTCAAAATTTAGATGATCTGAAGAGCAAGTTGCTATTGCTCAAT 1560  
Db 1501 ACTTGGCAAAACCAAAATCTCAAAATTTAGATGATCTGAAGAGCAAGTTGCTATTGCTCAAT 1560  
QY 1561 AGCTGATAAGTATACAAAGTCAAGTGGTTACATTTTGTGATGAACATGATATAACAGTGA 1620  
Db 1561 AGCTGATAAGTATACAAAGTCAAGTGGTTACATTTTGTGATGAACATGATATAACAGTGA 1620  
QY 1621 TGAAGGATGATATGTTAAGGCTCATATGGGCTATGCTGGATTTGGAAGAAATAG 1680  
Db 1621 TGAAGGATGATATGTTAAGGCTCATATGGGCTATGCTGGATTTGGAAGAAATAG 1680  
QY 1681 CCTTCTGATAAGCAAAAGTTCAGCTCAAGCTATACTAAAGAAAAAGTATCCTTACC 1740  
Db 1681 CCTTCTGATAAGCAAAAGTTCAGCTCAAGCTATACTAAAGAAAAAGTATCCTTACC 1740  
QY 1741 TCCATCTCCAGACGCGAGTGTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800  
Db 1741 TCCATCTCCAGACGCGAGTGTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800  
QY 1801 CAATCTGTGTAAGGGGAAAAAGCAATTCACCTGCTTCCATATATATGTTGAGCA 1860  
Db 1801 CAATCTGTGTAAGGGGAAAAAGCAATTCACCTGCTTCCATATATATGTTGAGCA 1860  
QY 1861 TACAGTTGAGTTTAAAAACGTTAAATTTGATTTCTCTATAAGGATCATTACCATATAT 1920  
Db 1861 TACAGTTGAGTTTAAAAACGTTAAATTTGATTTCTCTATAAGGATCATTACCATATAT 1920  
QY 1921 TAAATTTGCTTGTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980  
Db 1921 TAAATTTGCTTGTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980  
QY 1981 TTTGTTTGGACGATTAAGTACTACCTAGAACACCTTGAAGACGCTCCACATTTCTAATGA 2040  
Db 1981 TTTGTTTGGACGATTAAGTACTACCTAGAACACCTTGAAGACGCTCCACATTTCTAATGA 2040  
QY 2041 TGATGGGCAATTCGCAAGTGAAGCATGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA 2100  
Db 2041 TGATGGGCAATTCGCAAGTGAAGCATGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA 2100  
QY 2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACCTGTGTAGCCAGCAAGTCCC 2160  
Db 2101 TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACCTGTGTAGCCAGCAAGTCCC 2160  
QY 2161 TCAAGTAGAGATGTAAGAGTGAAGCCCACTCAAGAGAGCAAGATTTGTTGTTGCGAA 2220  
Db 2161 TCAAGTAGAGATGTAAGAGTGAAGCCCACTCAAGAGAGCAAGATTTGTTGTTGCGAA 2220  
QY 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAAATCTTAGCTGGTTTACGAATTA 2280  
Db 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAAATCTTAGCTGGTTTACGAATTA 2280  
QY 2281 TTTGATCTTCAAAATTTAGGATAACAAATAGTATCATGGCAGAGCAAGAAATTTACTTGC 2340  
Db 2281 TTTGATCTTCAAAATTTAGGATAACAAATAGTATCATGGCAGAGCAAGAAATTTACTTGC 2340

1261 TGTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGCTGACCAAGATTTTA 1320  
Db |  
1261 TGTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGCTGACCAAGATTTTA 1320  
Qy |  
1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380  
Db |  
1321 TGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGGTCG 1380  
Qy |  
1381 TAAATTCGTATTTCCAAAGCCTTAGACAAATATTAGAAGCCTTGAATGATGAATCGACTAA 1440  
Db |  
1381 TAAATTCGTATTTCCAAAGCCTTAGACAAATATTAGAAGCCTTGAATGATGAATCGACTAA 1440  
Qy |  
1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCAATTTAGCAACATTTACCCATCCAGAGG 1500  
Db |  
1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCAATTTAGCAACATTTACCCATCCAGAGG 1500  
Qy |  
1501 ACTTCGCAAAACCAATTTCTCAATTTGAGTATATCTGAAGACGAATTTGTTATTCCTCAAT 1560  
Db |  
1501 ACTTCGCAAAACCAATTTCTCAATTTGAGTATATCTGAAGACGAATTTGTTATTCCTCAAT 1560  
Qy |  
1561 AGCTGATAGTATACAAGCTCAGATGTTTACATTTTGTGAAACATGATATATCAGTGA 1620  
Db |  
1561 AGCTGATAGTATACAAGCTCAGATGTTTACATTTTGTGAAACATGATATATCAGTGA 1620  
Qy |  
1621 TGAAGGAGTATGATATGTAAGCCTCATATGGCCATAGTCTGATTTGGAAGAGATAG 1680  
Db |  
1621 TGAAGGAGTATGATATGTAAGCCTCATATGGCCATAGTCTGATTTGGAAGAGATAG 1680  
Qy |  
1681 CTTTCTGATAGGAAAAAGTTGAGCTCAAGCTATATCTAAGAAAAAGGTATCTTACC 1740  
Db |  
1681 CTTTCTGATAGGAAAAAGTTGAGCTCAAGCTATATCTAAGAAAAAGGTATCTTACC 1740  
Qy |  
1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800  
Db |  
1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800  
Qy |  
1801 CAATCTGTTGAAGGGAAAAAGATTTCCACTGTTCCACTCCATATATGTTGAGCA 1860  
Db |  
1801 CAATCTGTTGAAGGGAAAAAGATTTCCACTGTTCCACTCCATATATGTTGAGCA 1860  
Qy |  
1861 TACAGTTGAGTTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATATAT 1920  
Db |  
1861 TACAGTTGAGTTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATATAT 1920  
Qy |  
1921 TAAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980  
Db |  
1921 TAAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980  
Qy |  
1981 TTTGTTTGGAGATTAAGTACTACCTAGACACACCTGAGCAAGCTCCACATTTCTAATGA 2040  
Db |  
1981 TTTGTTTGGAGATTAAGTACTACCTAGACACACCTGAGCAAGCTCCACATTTCTAATGA 2040  
Qy |  
2041 TGGATGGGCAATGCGAGTGGATGATGTTTGGCAAGAAAGACACAGTGAAGATCCAA 2100  
Db |  
2041 TGGATGGGCAATGCGAGTGGATGATGTTTGGCAAGAAAGACACAGTGAAGATCCAA 2100  
Qy |  
2101 TAAGAACTTCAAGCGGATGAAGCGGATGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
Db |  
2101 TAAGAACTTCAAGCGGATGAAGCGGATGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
Qy |  
2161 TCAAGTAGAGACTGAAAGAGTGAAGCCCACTCAAGAGAGCAAGCTTTGTTGCGAA 2220  
Db |  
2161 TCAAGTAGAGACTGAAAGAGTGAAGCCCACTCAAGAGAGCAAGCTTTGTTGCGAA 2220  
Qy |  
2221 AGTAAAGGATTTAGTCTGAAGCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280  
Db |  
2221 AGTAAAGGATTTAGTCTGAAGCAATGCAACAGAACTCTAGCTGGTTTACGAATAA 2280  
Qy |  
2281 TTTGACTCTTCAAAATTTGATTAACATAGTATCATGGAGAGCAGAAAAATTTACTTGC 2340  
Db |  
2281 TTTGACTCTTCAAAATTTGATTAACATAGTATCATGGAGAGCAGAAAAATTTACTTGC 2340

2341 GTTGTAAAGGAGTAATCTCTTCTATCTGTAGTAAGGAAAAATAAAC 2389  
Db |  
2341 GTTGTAAAGGAGTAATCTCTTCTATCTGTAGTAAGGAAAAATAAAC 2389  
Qy |  
RESULT 2  
ABQ84819  
ID ABQ84819 standard; DNA; 2389 BP.  
XX  
AC ABQ84819;  
XX  
DT 04-SEP-2002 (first entry)  
XX  
S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.  
XX  
Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
KW antibacterial; Streptococcal infection; detection; gene; ds.  
XX  
Streptococcus pneumoniae.  
XX  
US2002061545-A1.  
XX  
23-MAY-2002.  
XX  
22-JAN-2001; 2001US-00765272.  
XX  
30-OCT-1997; 97US-00961083.  
XX  
(CHOI/) CHOI G H.  
PA (KUNS/) KUNSCH C A.  
PA (BARA/) BARASH S C.  
PA (DILL/) DILLON P J.  
PA (DOUG/) DOUGHERTY B.  
PA (FANN/) FANNON M R.  
PA (ROSE/) ROSEN C A.  
XX  
Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
Rosen CA;  
XX  
WPI; 2002-479261/51.  
XX P-PSDB; ABP54584.  
XX  
New Streptococcus pneumoniae antigens, useful for detecting Streptococcus  
and for preventing or attenuating disease caused by Streptococcus  
infection.  
XX  
Claim 1; Page 27; 70pp; English.  
XX  
ABQ84792 to ABQ84904 represents nucleic acids which encode the  
Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.  
pneumoniae antigens have antibacterial activity and can be used in  
vaccines. The S. pneumoniae antigens can also be used to prevent or  
attenuate a Streptococcal infection in an animal. The polynucleotides  
encoding the S. pneumoniae antigens can be used to detect Streptococcus  
nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning  
of S. pneumoniae ORFs (open reading frames) which are used in an example  
from the present invention  
XX  
SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;  
Query Match 100.0%; Score 2389; DB 6; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTTAAAGAAAAATAATCGTTTCTCTA 60  
Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTTAAAGAAAAATAATCGTTTCTCTA 60  
Qy 61 TATAGATGGAACAAACAGCGACGCAAAACCGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120  
Db 61 TATAGATGGAACAAACAGCGACGCAAAACCGGAGAAATTTGACTCTCTGATGAGGTTAGCAA 120  
Qy 121 GCGTGAAGGAATCAATGCTGAGCAATTCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180



AAV27351  
 ID AAV27351 standard; DNA; 2389 BP.  
 XX  
 AC AAV27351;  
 XX  
 DT 02-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae SP0036 nucleotide.  
 XX  
 DE Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 XX  
 KW detection; pneumonia; otitis media; meningitis; ss.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PH Key Location/Qualifiers  
 FT 2. 2389  
 FT CDS  
 FT /\*tag= a  
 FT /product= "SP0036"  
 FT /transl\_except= (pos:1367..1369,aa:Xaa)  
 FT /note= "no stop codon given; Xaa is unspecified"  
 XX  
 PN WO9818930-A2.  
 XX  
 XX 07-MAY-1998.  
 XX  
 XX 30-OCT-1997; 97WO-US019422.  
 XX  
 XX 31-OCT-1996; 96US-0029960P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;  
 XX  
 XX WPI; 1998-272224/24.  
 XX  
 XX P-PSDB; AAW55090.  
 XX  
 XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae  
 XX - or their epitope-containing fragments, useful in protective or  
 XX therapeutic vaccines, and for diagnosis.  
 XX  
 XX Claim 1; Page 59; 118pp; English.  
 XX  
 XX The present sequence encodes a protein from Streptococcus pneumoniae. The  
 XX nucleic acid sequence encoding the Streptococcus pneumoniae protein can  
 XX be useful in vaccines for inducing protective antibodies against  
 XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 XX are used to detect Streptococcus infection (by usual hybridisation or  
 XX amplification methods), also for isolating Streptococcus genes or their  
 XX allelic variants. The protein can be used similarly to detect specific  
 XX antibodies in standard immunoassays, especially for diagnosing or  
 XX monitoring infections. Antibodies which bind the protein are used to  
 XX detect corresponding antigens, to purify the protein and for passive  
 XX immunisation (optionally coupled to a toxin). Vaccines are administered,  
 XX e.g. by injection, orally or through the skin, typically at 0.01-1000  
 XX (especially 10-300) mu g/ml per dose  
 XX  
 XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;  
 XX  
 XX Query Match 100.0%; Score 2389; DB 2; Length 2389;  
 XX Best Local Similarity 100.0%; Pred. No. 0;  
 XX Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TTCTTAGAGTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60  
 Db 1 TTCTTAGAGTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTA 60  
 QY 61 TATAGATGGAAAAAAGCGACGCAAAACGAGAGATTTTCACTCCTGATGAGGTAGCAA 120  
 Db 61 TATAGATGGAAAAAAGCGACGCAAAACGAGAGATTTTCACTCCTGATGAGGTAGCAA 120  
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC 180  
 Db 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC 180

Db 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGTAC 180  
 QY 181 TTCATATGGCGACCACTATCATTTATTACATGGTAAGGTTTCCTTATGACGCTATCATCAG 240  
 Db 181 TTCATATGGCGACCACTATCATTTATTACATGGTAAGGTTTCCTTATGACGCTATCATCAG 240  
 QY 241 TGAAGAAATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300  
 Db 241 TGAAGAAATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300  
 QY 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360  
 Db 301 GGTCAAGGGTGGATATGTTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360  
 QY 361 TGCCACGCGGATTAACGTCGTTACAAAGAGGAAATCAATTCGACAAACAAAGAGCATAG 420  
 Db 361 TGCCACGCGGATTAACGTCGTTACAAAGAGGAAATCAATTCGACAAACAAAGAGCATAG 420  
 QY 421 TCACATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTTGCCTTGGCAGGTTGCGA 480  
 Db 421 TCACATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTTGCCTTGGCAGGTTGCGA 480  
 QY 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540  
 Db 481 AGGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540  
 QY 541 TGGTATGCTTATATCGTTCCTCATGAGATCAATTAACATTACATTCCTAAGAATGAGTT 600  
 Db 541 TGGTATGCTTATATCGTTCCTCATGAGATCAATTAACATTACATTCCTAAGAATGAGTT 600  
 QY 601 ATCAGCTAGCGAGTGGCTGCTGCAGAACGCTTCTATCTGGTGGAGGAAATCTGTCAA 660  
 Db 601 ATCAGCTAGCGAGTGGCTGCTGCAGAACGCTTCTATCTGGTGGAGGAAATCTGTCAA 660  
 QY 661 TTCAAGAACTATCGCGCAAAATAGCGATAACACTTCAAGAACAAATCTGGGTACTTTC 720  
 Db 661 TTCAAGAACTATCGCGCAAAATAGCGATAACACTTCAAGAACAAATCTGGGTACTTTC 720  
 QY 721 TGTAAAGCAATCCAGGAATCAAAATACAACTAACAAGCAACAGCAACAACTAACAGTCA 780  
 Db 721 TGTAAAGCAATCCAGGAATCAAAATACAACTAACAAGCAACAGCAACAACTAACAGTCA 780  
 QY 781 AGCAAGTCAAGTAAATGACATGATAGTCTCTTGAACAGCTCTCAAAATCGCTTTGAG 840  
 Db 781 AGCAAGTCAAGTAAATGACATGATAGTCTCTTGAACAGCTCTCAAAATCGCTTTGAG 840  
 QY 841 TCAACGACATGTAGAAATCTGATGGCCCTTGTCTTGTATCCAGCAAAATCAAGTCCGAAC 900  
 Db 841 TCAACGACATGTAGAAATCTGATGGCCCTTGTCTTGTATCCAGCAAAATCAAGTCCGAAC 900  
 QY 901 AGCTAGAGTGTGAGTGGCAGACGAGATCATTTACCACTTCATCCCTTACTCTCAAAAT 960  
 Db 901 AGCTAGAGTGTGAGTGGCAGACGAGATCATTTACCACTTCATCCCTTACTCTCAAAAT 960  
 QY 961 GTCTGAATTTGGAAGAACGAATCGCTGTTATTTCCCTTCTGTTATCGTTTCAAAACATTG 1020  
 Db 961 GTCTGAATTTGGAAGAACGAATCGCTGTTATTTCCCTTCTGTTATCGTTTCAAAACATTG 1020  
 QY 1021 GGTACCAAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG 1080  
 Db 1021 GGTACCAAGATTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG 1080  
 QY 1081 CCGGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGT 1140  
 Db 1081 CCGGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGT 1140  
 QY 1141 ACMAAAGTTGGGAGGATATGTTTGAAGAAAGGAGCATCTCTCGTTATGCTTTTC 1200  
 Db 1141 ACMAAAGTTGGGAGGATATGTTTGAAGAAAGGAGCATCTCTCGTTATGCTTTTC 1200  
 QY 1201 GAAAGATTTACCATCTGAAACTGTTAAATAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260  
 Db 1201 GAAAGATTTACCATCTGAAACTGTTAAATAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 20:07:08 ; Search time 915 Seconds  
(without alignments)  
11091.743 Million cell updates/sec

Title: US-09-765-271-55  
Perfect score: 2389  
Sequence: 1 TTCTTACGAGTTGGAGTCTG.....TAAGTAAGCAAAATAAATAAC 2389

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : N Geneseq 23Jan04:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001s:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002s:\*
- 7: geneseqn2003as:\*
- 8: geneseqn2003bs:\*
- 9: geneseqn2003cs:\*
- 10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2389	100.0	2389	2	AAV27351 Streptococcus
2	2389	100.0	2389	6	ABQ84819 S. pneumoniae
3	2389	100.0	2389	9	ADC45136 S. pneumoniae
4	2389	100.0	2451	3	AAA47604 Recombinant
5	2338	97.9	2406	7	ABX06886 S. pneumoniae
6	2338	97.9	8195	2	AAV52227 Streptococcus
7	2338	97.9	110000	7	ABX56454_09 Continuation (10 o
8	2338	97.9	110000	7	ABX56454_10 Continuation (11 o
9	148	6.2	2523	3	AAA65731 Streptococcus
10	148	6.2	2647	3	AAA65736 Streptococcus
11	148	6.2	2647	6	ABK15103 DNA encod
12	53	2.2	504	3	AAA08556 S. pneumoniae
13	53	2.2	2163	2	AAV25394 Streptococcus
14	53	2.2	2290	6	AAV27356 Streptococcus
15	53	2.2	2290	6	ABQ84824 S. pneumoniae
16	53	2.2	2290	9	ADC45146 S. pneumoniae
17	53	2.2	2359	2	AAV52376 Streptococcus
18	53	2.2	2457	7	ABX06885 S. pneumoniae
19	53	2.2	2478	3	AAA08557 S. pneumoniae
20	53	2.2	2481	3	AAA05417 Streptococcus
21	53	2.2	2517	7	ABX06705 S. pneumoniae
22	53	2.2	2531	3	AAA47605 Recombinant
23	53	2.2	2531	3	AAA47602 Recombinant

24	53	2.2	2639	3	AAA65737 Streptococcus
25	53	2.2	2639	7	ABK15104 DNA encod
26	53	2.2	110000	6	Continuation (9 of
27	50	2.1	492	2	AAV25393 Streptococcus
28	44	1.8	973	2	AAV52488 Streptococcus
29	38	1.6	3171	3	AAA65739 Streptococcus
30	30	1.3	40	2	AAV27492 Streptococcus
31	30	1.3	40	3	AAA47601 Primer fo
32	30	1.3	40	6	ABQ84960 Streptococcus
33	30	1.3	40	9	ADC45363 S. pneumoniae
34	27	1.1	36	3	AAA47599 Primer fo
35	27	1.1	37	2	AAV27491 Streptococcus
36	27	1.1	37	6	ABQ84959 Streptococcus
37	27	1.1	37	9	ADC45362 S. pneumoniae
38	26	1.1	35	3	AAA47600 Primer fo
39	25	1.0	33	3	AAA65762 Streptococcus
40	25	1.0	33	6	ABK33684 S. pneumoniae
41	25	1.0	34	3	AAA65764 Streptococcus
42	25	1.0	34	6	ABK33686 S. pneumoniae
43	25	1.0	35	6	ABK33719 S. pneumoniae
44	24	1.0	1342	2	AAV27414 Streptococcus
45	24	1.0	1342	6	ABQ84882 S. pneumoniae
46	24	1.0	1342	9	ADC45262 S. pneumoniae
47	24	1.0	1398	6	ABN66839 Streptococcus
48	24	1.0	1455	3	AAA05473 Streptococcus
49	24	1.0	1455	3	AAZ91804 Streptococcus
50	24	1.0	1455	3	AAA65733 Streptococcus
51	24	1.0	1455	3	AAA47603 Recombinant
52	24	1.0	2528	3	AAA65738 Streptococcus
53	24	1.0	2528	6	ABK15105 DNA encod
54	24	1.0	3117	7	ABX06706 S. pneumoniae
55	24	1.0	3120	3	AAA65730 Streptococcus
56	24	1.0	5048	3	AAA65735 Streptococcus
57	24	1.0	5048	6	ABK15101 DNA encod
58	24	1.0	6867	2	AAV52325 Streptococcus
59	24	1.0	110000	6	ABN71527_18 Continuation (19 o
60	23	0.9	77287	8	AAD58261 Murine tu
61	23	1.0	167739	8	AAD58258 Murine tu
62	22	0.9	1547	3	AAV37581 Arabidops
63	21	0.9	29	6	AAA65761 Streptococcus
64	21	0.9	29	6	ABK33683 S. pneumoniae
65	21	0.9	690	8	ACL22740 DNA clone
66	21	0.9	110000	7	ABQ84281_2 Continuation (3 of
67	20	0.8	501	6	AAI69344 C. albica
68	20	0.8	1146	6	AAA05814 Group B S
69	20	0.8	2466	6	ABN69535 Streptococcus
70	20	0.8	2466	6	ABN70334 Streptococcus
71	20	0.8	2469	3	AAA05811 Group B S
72	20	0.8	2469	3	AAA65740 Streptococcus
73	20	0.8	2472	5	AAA00038 Streptococcus
74	20	0.8	2472	3	AAA65741 Streptococcus
75	20	0.8	2478	6	ABN66838 Streptococcus
76	20	0.8	2478	5	AAA00036 Streptococcus
77	20	0.8	5158	6	AAI69348 C. albica
78	20	0.8	5215	6	AAV91105 Group B S
79	20	0.8	110000	6	ABN71527_12 Continuation (13 o
80	19	0.8	164	7	ABX54820 Bovine ES
81	19	0.8	426	5	ABV02215 Human pro
82	19	0.8	427	4	AAV30789 Human CDN
83	19	0.8	468	5	ABV11384 Human pro
84	19	0.8	481	5	ABV32529 Human pro
85	19	0.8	481	5	ABV41454 Human pro
86	19	0.8	693	6	ABQ89355 Human pro
87	19	0.8	693	7	ADB82298 Human CDN
88	19	0.8	1020	7	ACA28893 Prokaryot
89	19	0.8	1203	5	ABV21401 Human pro
90	19	0.8	1205	5	ABV27219 Human pro

ALIGNMENTS

RESULT 1



LLATVYVHPNPHSDNGFGNASDHVORKNQADTNOTKPSKPKPOTKPEEE  
TPREKQSEKPSKPKTEPESEPESEEPQVETKVEKLRRAEDLLGKIQDPIIK  
SNAKETITGLKNLLFGTDNNITMAEAKLLALKESK"

## ORIGIN

Query Match 2.2%; Score 53; DB 12; Length 2166;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 1035  
Db 688 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 740

RESULT 23  
AR120270 AR120270 2290 bp DNA linear PAT 16-MAY-2001  
LOCUS Sequence 65 from patent US 6159469.  
DEFINITION  
ACCESSION AR120270  
VERSION AR120270.1 GI:14103846  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2290)  
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,  
Fannon, M.R. and Rosen, C.A.  
TITLE Streptococcus pneumoniae antigens and vaccines  
JOURNAL Patent: US 6159469-A 65 12-DEC-2000;  
FEATURES Location/Qualifiers  
source 1..2290  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

Query Match 2.2%; Score 53; DB 6; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 1035  
Db 944 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 996

RESULT 24  
AR340961 AR340961 2290 bp DNA linear PAT 17-AUG-2003  
LOCUS Sequence 65 from patent US 6573082.  
DEFINITION  
ACCESSION AR340961  
VERSION AR340961.1 GI:33732940  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2290)  
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,  
Fannon, M.R. and Rosen, C.A.  
TITLE Streptococcus pneumoniae antigens and vaccines  
JOURNAL Patent: US 6573082-A 65 03-JUN-2003;  
FEATURES Location/Qualifiers  
source 1..2290  
/organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

Query Match 2.2%; Score 53; DB 6; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 1035  
Db 944 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 996

## RESULT 25

BD063279 BD063279 2290 bp DNA linear PAT 27-AUG-2002  
LOCUS Streptococcus pneumoniae antigens and vaccines.  
DEFINITION  
ACCESSION BD063279  
VERSION BD063279.1 GI:22608882  
KEYWORDS JP 2001505415-A/33.  
SOURCE unidentified  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 2290)  
AUTHORS Kunsch, C.A., Choi, G.H., Johnson, S.L. and Hromockyj, A.  
TITLE Streptococcus pneumoniae antigens and vaccines  
JOURNAL Patent: JP 2001505415-A 33 24-APR-2001;  
HUMAN GENOME SCIENCES INC  
COMMENT PN JP 2001505415-A/33  
PD 24-APR-2001  
PF 30-OCT-1997 JP 1998520667  
PR 31-OCT-1996 US 60/029960  
PI CHARLES A KUNSCH, GIL H CHOI, SYDNOR L JOHNSON, ALEX HROMOCKYJ PC  
C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC  
G01N33/569,  
PC G01N33/68  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers.

FEATURES  
source 1..2290  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 2.2%; Score 53; DB 6; Length 2290;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 983 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 1035  
Db 944 GCTCGTATTATTCCTTCGTTATCGTTCAAAACCATTCGGTACCAGATTCAG 996

Search completed: October 1, 2004, 01:51:32  
Job time : 9285 secs



[illegible]



Db 18433 TTTGAAATAAGGTCGTAATCTGATTTCAASCCTTGGCAAAATTTTAGAACGCTTG 18492  
 QY 1424 AATGATGAATCGAC -TAAATAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACC 1482  
 Db 18493 AATGATGAATCGACTTAAATAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCACC 18552  
 QY 1483 AATTACCATTGAGGAGCTTGGCAACCAAAATTTCTAAATTTGAGTATCTAGAGAGCA 1542  
 Db 18553 AATTACCATTGAGGAGCTTGGCAACCAAAATTTCTAAATTTGAGTATCTAGAGAGCA 18612  
 QY 1543 AGTTCGTAATTCGTAATTTAGCTGTAAGTATATAACAACGTCAGATGGTATACATTTTGGATGA 1602  
 Db 18613 AGTTCGTAATTCGTAATTTAGCTGTAAGTATATAACAACGTCAGATGGTATACATTTTGGATGA 18672  
 QY 1603 ACATGATATAATTCAGTGAATGAAGAGATGCTATATGTAACGCTCATATGGGCTATAGTCA 1662  
 Db 18673 ACATGATATAATTCAGTGAATGAAGAGATGCTATATGTAACGCTCATATGGGCTATAGTCA 18732  
 QY 1663 CTGGATTGGAAAAGATAGCTTTCTGATAAGAAAAAGTTGCAGCTCAAGCCTATATAA 1722  
 Db 18733 CTGGATTGGAAAAGATAGCTTTCTGATAAGAAAAAGTTGCAGCTCAAGCCTATATAA 18792  
 QY 1723 AGAAAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGA 1782  
 Db 18793 AGAAAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGA 18852  
 QY 1783 TAGTCAGCAGCTATTTTCAATTCGTTGAAAGGGGAAAAAGAAATTCCTGTCGACT 1842  
 Db 18853 TAGTCAGCAGCTATTTTCAATTCGTTGAAAGGGGAAAAAGAAATTCCTGTCGACT 18912  
 QY 1843 TCATATATGTTGAGCATACAGTTGAGGTTAAAGCGTAATTTGATTTATTCCTCATAA 1902  
 Db 18913 TCATATATGTTGAGCATACAGTTGAGGTTAAAGCGTAATTTGATTTATTCCTCATAA 18972  
 QY 1903 GGATCATTTACCAATAATTTAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAA 1962  
 Db 18973 GGATCATTTACCAATAATTTAAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAA 19032  
 QY 1963 TGGCTATACCTTGGAAAGATTTGTTGCGACGATTAAGTACTACGTAGAAACACCTGACGA 2022  
 Db 19033 TGGCTATACCTTGGAAAGATTTGTTGCGACGATTAAGTACTACGTAGAAACACCTGACGA 19092  
 QY 2023 AGTCCACATTTCTAATGATGGATGGGCAATCCAGTGAAGCATGCTTGAAGCAAGAAAGCA 2082  
 Db 19093 AGTCCACATTTCTAATGATGGATGGGCAATCCAGTGAAGCATGCTTGAAGCAAGAAAGCA 19152  
 QY 2083 CCACAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAAGCCAGTAGAGAAACACCC 2142  
 Db 19153 CCACAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAAGCCAGTAGAGAAACACCC 19212  
 QY 2143 TGTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAGATGAAGCCCACTCAAGAGCC 2202  
 Db 19213 TGTGAGCCAGAGTCCCTCAAGTAGAGACTGAAAGATGAAGCCCACTCAAGAGCC 19272  
 QY 2203 AGAAGTTTTCGTTGCGAAGTAAAGTCTAGTCTGAAAGCCAAATGCAACGAACTCT 2262  
 Db 19273 AGAAGTTTTCGTTGCGAAGTAAAGTCTAGTCTGAAAGCCAAATGCAACGAACTCT 19332  
 QY 2263 AGCTGTTTACGAAATAATTTGACTCTTCAAAATTTATGGATAACAATGATATCATGCGAGA 2322  
 Db 19333 AGCTGTTTACGAAATAATTTGACTCTTCAAAATTTATGGATAACAATGATATCATGCGAGA 19392  
 QY 2323 AGCAGAAAAATTTACTTCGCTGTTTAAAGAAAGTAACTCTTCTGTAAGTAAAGAAA 2382  
 Db 19393 AGCAGAAAAATTTACTTCGCTGTTTAAAGAAAGTAACTCTTCTGTAAGTAAAGAAA 19452  
 QY 2383 AATAAAC 2389  
 Db 19453 AATAAAC 19459

RESULT 14

SPNEU1929 702 bp DNA linear HTG 11-JUL-2001  
 LOCUS Streptococcus pneumoniae clone G54, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION AL449951 GI:11545176  
 ACCESSION HTG: HTGS PHASE2  
 VERSION Streptococcus pneumoniae  
 KEYWORDS Streptococcus pneumoniae  
 SOURCE Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 ORGANISM Streptococcus.  
 REFERENCE 1  
 AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,  
 Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de  
 Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.,  
 and Garcia-Bustos, J.F.  
 TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae  
 JOURNAL type 19f clinical isolate  
 MEDLINE Microb. Drug Resist. 7 (2), 99-125 (2001)  
 PUBMED 21335329  
 REFERENCE 11442348  
 AUTHORS 2 (bases 1 to 702)  
 Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,  
 Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de  
 Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and  
 Garcia-Bustos, J.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,  
 Severo Ochoa 2, 28760 Tres Cantos, SPAIN  
 COMMENT \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES  
 Location/Qualifiers  
 1..702  
 /organism="Streptococcus pneumoniae"  
 /mol\_type="genomic DNA"  
 /serotype="19F"  
 /db\_xref="taxon:1313"  
 /clone="G54"  
 ORIGIN  
 Query Match 13.6%; Score 326; DB 2; Length 702;  
 Best Local Similarity 99.1%; Pred. No. 9,7e-164;  
 Matches 576; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 265 AACTATAAGCTAAAGATGAGGATATTGTTAATGAGGTCAAGGTGATATGTTATCAAG 325  
 Db 48 AACTATAAGCTAAAGATGAGGATATTGTTAATGAGGTCAAGGTGATATGTTATCAAG 107  
 QY 326 GTAGATGGAAAAATCTATGTTTACCTTAAGGATGCTGCCACCGGATAACGTCGGTACA 385  
 Db 108 GTAGATGGAAAAATCTATGTTTACCTTAAGGATGCTGCCACCGGATAACGTCGGTACA 167  
 QY 386 AAGAGGAAATCATCGACAAAACAAAGACATAGTCAACATCGTGAAGTGAACCTCCA 445  
 Db 168 AAGAGGAAATCAATCGACAAAACAAAGACATAGTCAACATCGTGAAGTGAACCTCCA 227  
 QY 446 AGAAACGATGGTCTGCTTGGCAGCGTTGCGAAGGAGCGTATATCAAGATGATGGT 505  
 Db 228 AGAAACGATGGTCTGCTTGGCAGCGTTGCGAAGGAGCGTATATCAAGATGATGGT 287  
 QY 506 TATATCTTTAAATGCTTTCTGATATCATAGAGGATATCGGTGATGCTTATATCGTTCCTCAT 565  
 Db 288 TATATCTTTAAATGCTTTCTGATATCATAGAGGATATCGGTGATGCTTATATCGTTCCTCAT 347  
 QY 566 GAGATCATTTACCATTTACATTCCTTAAGATGAGTATCGCTAGCGAGTGGCTGCTGCA 625  
 Db 348 GAGATCATTTACCATTTACATTCCTTAAGATGAGTATCGCTAGCGAGTGGCTGCTGCA 407  
 QY 626 GAAGCCTTCCTATCTGCTCGAGGAATCTCTCAAAATTCAGAACTATCCCGCAAAAAT 685  
 Db 408 GAAGCCTTCCTATCTGCTCGAGGAATCTCTCAAAATTCAGAACTATCCCGCAAAAAT 467



Db 6428 TAATCTGATTTCCAGCCTTACAGAAATTAATTAGAACGCTTGAATGATGAACTCGACTAA 6369  
 QY 1441 TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCAGCAATTTACCAATCCAGAGCG 1500  
 Db 6368 TAAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCAGCAATTTACCAATCCAGAGCG 6309  
 QY 1501 ACTTGGCAACCAAAATTTCTCAATTTGAGTATCTAGAGCAAGTTCGTTATTTGCTCAAT 1560  
 Db 6308 ACTTGGCAACCAAAATTTCTCAATTTGAGTATCTAGAGCAAGTTCGTTATTTGCTCAAT 6249  
 QY 1561 ACCTGATAGTATACAACTGAGTGGTTACATTTTGTAGAACATGATATATATCACTGCA 1620  
 Db 6248 ACCTGATAGTATACAACTGAGTGGTTACATTTTGTAGAACATGATATATATCACTGCA 6189  
 QY 1621 TGAAGGAGATGATATGTAACCCCTCATATGGCCATAGTCACTGGATTTGGAAGAGATAG 1680  
 Db 6188 TGAAGGAGATGATATGTAACCCCTCATATGGCCATAGTCACTGGATTTGGAAGAGAGATAG 6129  
 QY 1681 CTTTCTGATAGGAAAAAGTTGCACTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 1740  
 Db 6128 CTTTCTGATAGGAAAAAGTTGCACTCAAGCCTATATCTAAAGAAAAAGGTATCTTACC 6069  
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATGATGAGCAGCTTATTTA 1800  
 Db 6068 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATGATGAGCAGCTTATTTA 6009  
 QY 1801 CAATCGTGTGAAGGGGAAAAAGCAATCCCACTGCTGCACTTCCATATATGTTGAGCA 1860  
 Db 6008 CAATCGTGTGAAGGGGAAAAAGCAATCCCACTGCTGCACTTCCATATATGTTGAGCA 5949  
 QY 1861 TACAGTTGAGTTAAAAACGGTAAATTTGATTTCTCATAGGATCAATACCAATATAT 1920  
 Db 5948 TACAGTTGAGTTAAAAACGGTAAATTTGATTTCTCATAGGATCAATACCAATATAT 5889  
 QY 1921 TAAATTTGCTTGTGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980  
 Db 5888 TAAATTTGCTTGTGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 5829  
 QY 1981 TTTGTTTGGCAGATTAAGTACTAGTACGTAAGACACCTGACCAAGCTCCACATCTCAATGA 2040  
 Db 5828 TTTGTTTGGCAGATTAAGTACTAGTACGTAAGACACCTGACCAAGCTCCACATCTCAATGA 5769  
 QY 2041 TGGATGGGCAATGCCAGTGAAGATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100  
 Db 5768 TGGATGGGCAATGCCAGTGAAGATGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 5709  
 QY 2101 TAAGAACTTCAAAGCGGATGAAGACGAGTACGTAAGACACCTGACGAGCAAGTCCC 2160  
 Db 5708 TAAGAACTTCAAAGCGGATGAAGACGAGTACGTAAGACACCTGACGAGCAAGTCCC 5649  
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAGAGCAAGTTCGTTGGCA 2220  
 Db 5648 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAGAGCAAGTTCGTTGGCA 5589  
 QY 2221 AGTAACGGATTCAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280  
 Db 5588 AGTAACGGATTCAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 5529  
 QY 2281 TTTGACTCTTCAAATTTATGATAACATAGTATCATGCGCAGAGCAAAAAATCTTGC 2340  
 Db 5528 TTTGACTCTTCAAATTTATGATAACATAGTATCATGCGCAGAGCAAAAAATCTTGC 5469  
 QY 2341 GTTGTAAAAGGAGTAATCTTCTATCTGTAAGTAAAGGAAAAATAAAC 2389  
 Db 5468 GTTGTAAAAGGAGTAATCTTCTATCTGTAAGTAAAGGAAAAATAAAC 5420

RESULT 13  
 SPNEU1915  
 LOCUS  
 DEFINITION  
 Streptococcus pneumoniae clone G54, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
 ACCESSION  
 AL449937.1  
 VERSION  
 GI:11545162

KEYWORDS HTG; HTGS PHASE2.  
 SOURCE Streptococcus pneumoniae  
 ORGANISM Streptococcus pneumoniae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1  
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,  
 Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de  
 Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.  
 and Garcia-Bustos,J.F.  
 Annotated draft genomic sequence from a Streptococcus pneumoniae  
 type 19F clinical isolate  
 Microb. Drug Resist. 7 (2), 99-125 (2001)  
 21353329  
 11442348  
 2 (bases 1 to 20035)  
 Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,  
 Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de  
 Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and  
 Garcia-Bustos,J.F.  
 Direct Submission  
 Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,  
 Severo Ochoa 2, 28760 Tres Cantos, SPAIN  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 Location/Qualifiers  
 1. 20035  
 /organism="Streptococcus pneumoniae"  
 /mol\_type="genomic DNA"  
 /serotype="19F"  
 /db\_xref="taxon:1313"  
 /clone="G54"  
 ORIGIN  
 Query Match 38.8%; Score 927; DB 2; Length 20035;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 1497; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 QY 884 CAAATCAAAAGTCGAAACAGCTAGAGGTGTCAGTGCACAGGATCATATACCATTTC 943  
 Db 17953 CAAATCAAAAGTCGAAACAGCTAGAGGTGTCAGTGCACAGGATCATATACCATTTC 18012  
 QY 944 ATCCCTTACTCTCAAAATGCTGAATTTGGAAGACGAATCGCTGATATTCCCTTCGT 1003  
 Db 18013 ATCCCTTACTCTCAAAATGCTGAATTTGGAAGACGAATCGCTGATATTCCCTTCGT 18072  
 QY 1004 TATCGTTCAAAACCATTTGGGTACCAAGATTCAAGCCAGAACCAAGTCCCAACCGACT 1063  
 Db 18073 TATCGTTCAAAACCATTTGGGTACCAAGATTCAAGCCAGAACCAAGTCCCAACCGACT 18132  
 QY 1064 CCGGAACCTAGTCCAGCGCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATTTCTCT 1123  
 Db 18133 CCGGAACCTAGTCCAGCGCCGCAACCTGCAACCAATCTTAAATAGACTCAAAATTTCTCT 18192  
 QY 1124 TTGTTAGTCACTGCTGTAAGAAAAGTTGGGAAGGATATGATTTCCGAAGAAAGGCATC 1183  
 Db 18193 TTGTTAGTCACTGCTGTAAGAAAAGTTGGGAAGGATATGATTTCCGAAGAAAGGCATC 18252  
 QY 1184 TCTCGTTATGCTTTGCGAAAGTTTACCATCTGAAACTGTTAAATACTTTGAAAGCAAG 1243  
 Db 18253 TCTCGTTATGCTTTGCGAAAGTTTACCATCTGAAACTGTTAAATACTTTGAAAGCAAG 18312  
 QY 1244 TTATCAAAAACAGAGAGTCTTTACACACTTTTAACTGCTTAAATAAGAAATCTTGCTCT 1303  
 Db 18313 TTATCAAAAACAGAGAGTCTTTACACACTTTTAACTGCTTAAATAAGAAATCTTGCTCT 18372  
 QY 1304 CTGACCAAGAAATTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTG 1363  
 Db 18373 CTGACCAAGAAATTTATGATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTG 18432  
 QY 1364 TTGGAATAAGGGTCTGAATTTCTGATTTCCAGCCTTAGACAAATTTAGAACGCTTG 1423

GVAVPHGDHYHIFYPYSOMSELEERIALIIPLYRSNWVPSRPEQSPQTPRPSPG PQAPNLIKIDSSNSLSQVLRKVGEGYVFEKGLSIRYVFAKDLFSEIVKNLESLSKQ ESVHTTAKVNDVAPRDOEFYKAYNLLTAHAKALFENKGRNSDFQALDKLERLND BSTNKEKLVALLAPLAPITPHERLGNKNSQIETEDVEVIAQADKYTTSDGYIFDE HD11SDGDAVYVTHMGSHWIKDLSDEKVAQAAYTEKGLILPPSPDADVKANPT GDSAAIYRNVKSGRIPLVRLPVVEHTVEKGNLIIPKOHVHNIKFADFDDHTY KAPNGYTLDELFIATIKYVHERPDERPHSNDGNGNASHVLGKDKHSDPNKNFKADEE PVEETPAPEVPEVQVETEKVQALKEAEVLLAKVDSLSUKANATETLAGLRNLLTIQIM DNNSIMAEKLLALLKGNPSPPSSKEKIN"	gene
/complement (8122. .9855) /gene="pts1" /note="synonym: spr1062" /complement (8122. .9855) /gene="pts1" /EC number="2.7.3.9" /codon_start=1 /transl_table=11 /label=spr1062 /product="PTS enzyme 1" /protein_id="AAK99866.1" /db_xref="GI:15458684" /translation="MTEMLKGIASDGVAVAKAYLLVQDLSFETITVEDTNAEBEARL DAALQASODELSVIREKXAVGTGEEAAQVFDHLVLPADPEMISQIKETIRAKKYNAAE AGLKVDMFIIITFGMEDNPMYQRAADIRDTKRVLANLLKGLPNPASINEEVIV IAHDLTPSDAQIDKNFYKAFVNIIGRTSHSAIMARTLEIAVLGTNNITETIVKQGD ILAVNGITGEVINPTDEQADEFKAAGEAYAKQAEWALLKDAQTVTADGKHFELAA IGTPKDEGVNNGAEVGLYRTEFLYMDSPQPTDEDEQTEYAKVLENGMKPVPVR TMDIGDELKPLFDMPHENPFLGFRLAISSETGDMFRTOIRALLRASVHGQIRI MFPVALLKEFAAKAAVDFEKNALLAEGVADNITQVIMIEIPAAAMLADOFKBEV DPFSGTNDLLOIYVMAADRMEQVSLYQVYNPILRLINNIKAHAEGKWMGCE MAGDOQAVPLLVMGLDPEFSMSATSVLRTSLMKKLDATAKMEEYANRALTEGCTMBEV LBLEKYEYNFD" complement (9861. .10124) /gene="ptsH" /note="synonym: spr1063" complement (9861. .10124) /gene="ptsH" /codon_start=1 /transl_table=11 /label=spr1063 /product="Histidine-containing phosphocarrier protein of the PTS" /protein_id="AAK99867.1" /db_xref="GI:15458685" /translation="MASKDPHVVAETGIIHARPATLLVQTRASKFASDITLEYKKSVINL KSTMGVMSLGVGGQADVTISAEAGDADDAIAA1SETMEKEGLA"	gene
Query Match 68.0%; Score 1624; DB 1; Length 10320; Best Local Similarity 99.4%; Pred. No. 0; Matches 2374; Conservative 0; Mismatches 15; Indels 0; Gaps 0;	CDS
QY 1 TTCTTACAGTGTGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 60 DB 7808 TTCTTACAGTGTGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA 7749	
QY 61 TATAGATGGAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAA 120 DB 7748 TATAGATGGAACAAACGACGCGCAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAA 7689	
QY 121 GCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180 DB 7688 GCGTGAAGGAATCAATCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 7629	
QY 181 TTCACATGGCCACCACTATCATTATTAACAATGGTAAGGTTCTCTTATGACGCTATCATCAG 240 DB 7628 TTCACATGGCCACCACTATCATTATTAACAATGGTAAGGTTCTCTTATGATGCGAATTTTCAG 7569	
QY 241 TGAAGAATTAATCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA 300 DB 7568 TGAAGAATTAATCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGA 7509	
QY 301 GGTCAAGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTTAAGGATGC 360	

7508	Db	GGTCAAGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTTAAGGATGC 7449
361	QY	TGCCACGCGATACAGTCCGTACAAAAGAGGAAATCAATGCACAAAACAAAGACATAG 420
7448	Db	TGCCACGCGATACAGTCCGTACAAAAGAGGAAATCAATGCACAAAACAAAGACATAG 7389
421	QY	TCAACATCGTGAAGGTGGAACCTCCAAAGAACGATGGTGTCTGCTTGGCTGGGACGTTGCGA 480
7388	Db	TCAACATCGTGAAGGTGGAACCTCCAAAGAACGATGGTGTCTGCTTGGCTGGGACGTTGCGA 7329
481	QY	AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
7328	Db	AGGACGCTATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 7269
541	QY	TGGTGTATGCTTATATCGTTCCTCATGGAGATCATTTACCATTACATTCCTTAAGAATGAGTT 600
7268	Db	TGGTGTATGCTTATATCGTTCCTCATGGAGATCATTTACCATTACATTCCTTAAGAATGAGTT 7209
601	QY	ATCAGCTAGCGAGTTGGCTGTGCGAAGCCTTCCTATCTGCTGAGGAAATCTGTCAA 660
7208	Db	ATCAGCTAGCGAGTTGGCTGTGCGAAGCCTTCCTATCTGCTGAGGAAATCTGTCAA 7149
661	QY	TTCAAGAACCTATCGCGACAAAATAGCGATACACCTTCAAGAACAACTGGGTACCTTC 720
7148	Db	TTCAAGAACCTATCGCGACAAAATAGCGATACACCTTCAAGAACAACTGGGTACCTTC 7089
721	QY	TGTAAAGCAATCCAGGAACCTACAAATACTAAACAAGCAACAACAGCAACACTAACAGTCA 780
7088	Db	TGTAAAGCAATCCAGGAACCTACAAATACTAAACAAGCAACAACAGCAACACTAACAGTCA 7029
781	QY	AGCAAGTCAAGGTAATGACATTTGATAGTCTCTTTGAAAACAGCTCTACAAATGCTTTGAG 840
7028	Db	AGCAAGTCAAGGTAATGACATTTGATAGTCTCTTTGAAAACAGCTCTACAAATGCTTTGAG 6969
841	QY	TCAACGACATGATAGATCTGATGGCTGTCTTTGATTCAGACCAAACTACAAAGTCGAAC 900
6968	Db	TCAACGACATGATAGATCTGATGGCTGTCTTTGATTCAGACCAAACTACAAAGTCGAAC 6909
901	QY	AGCTAGAGGTGTGTCAGTGCACACGAGAGATCATTTACCCTTCATCCCTTACTCTCAAT 960
6908	Db	AGCTAGAGGTGTGTCAGTGCACACGAGAGATCATTTACCCTTCATCCCTTACTCTCAAT 6849
961	QY	GTCTGAATTTGGAAGAACGAATCGCTGTATATTCCTCTTCCTGTTATCGTTCAACCACTG 1020
6848	Db	GTCTGAATTTGGAAGAACGAATCGCTGTATATTCCTCTTCCTGTTATCGTTCAACCACTG 6789
1021	QY	GGTACCGAGATCAAGGCGCAACCAACCACTCCACACCGACTCCGGAACCTAGTCCAGG 1080
6788	Db	GGTACCGAGATCAAGGCGCAACCAACCACTCCACACCGACTCCGGAACCTAGTCCAGG 6729
1081	QY	CCGCAACCTGACCAAAATCTTAAATAGACTCAAATTCCTTTGGTTAGTCAAGTGGT 1140
6728	Db	CCGCAACCTGACCAAAATCTTAAATAGACTCAAATTCCTTTGGTTAGTCAAGTGGT 6669
1141	QY	ACGAAAAAGTTGGGGAAGGATATGATTTTCGAAGAAAGAGGCTATCTCGTTATGCTTTGCG 1200
6668	Db	ACGAAAAAGTTGGGGAAGGATATGATTTTCGAAGAAAGAGGCTATCTCGTTATGCTTTGCG 6609
1201	QY	GAAAGATTTTACCATCTGAAACTGTAAAAATCTTGAAGCAAGTATCAAAAACAAAGAGAG 1260
6608	Db	GAAAGATTTTACCATCTGAAACTGTAAAAATCTTGAAGCAAGTATCAAAAACAAAGAGAG 6549
1261	QY	TGTTTTCACACATTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAAATTTTA 1320
6548	Db	TGTTTTCACACATTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAAATTTTA 6489
1321	QY	TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTGTTTGNAAATAAGGGTCG 1380
6488	Db	TGATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTGTTTGNAAATAAGGGTCG 6429
1381	QY	TAAATTCGATTTCCAAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGATGACATAA 1440

RESULT 12	AE008479	10320 bp	DNA	linear	BCT 13-SEP-2001
LOCUS	Streptococcus pneumoniae R6	section 95 of 184	of the complete genome.		
DEFINITION	AE008479 AE007317				
ACCESSION	AE008479.1	GI:15458677			
KEYWORDS					
SOURCE	Streptococcus pneumoniae R6				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
REFERENCE	1 (bases 1 to 10320)				
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., LaGace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicolas, T.I., Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rokey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				
TITLE	Genome of the bacterium Streptococcus pneumoniae strain R6				
JOURNAL	J. Bacteriol. 183 (19), 5709-5717 (2001)				
MEDLINE	21429245				
PUBMED	11542334				
REFERENCE	2 (bases 1 to 10320)				
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., LaGace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicolas, T.I., Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rokey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				
Direct Submission					
Submitted (27-JUN-2001)	Infectious Diseases Research, Eli Lilly and Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA				
Location/Qualifiers	1. .10320				
organism="Streptococcus pneumoniae R6"					
/mol_type="genomic DNA"					
/strain="R6"					
/db_xref="taxon:171101"					
complement (91. .738)					
/gene="spr1056"					
complement (91. .738)					
/gene="spr1056"					
/codon_start=1					
/transl_table=11					
/label=spr1056					
/product="Hypothetical protein"					
/protein_id="AAK99860.1"					
/db_xref="GI:15458678"					
/translation="MTYKKNITIEYEMSEVDNEAVNVTSTQCLANSIVQSLFKFDFKNGSLDIFTKPSLVKEGDIYGRDSQVDNLRVIGNIPNYLTNRKXSLNMNRGCMGDFDFDIYDLHVAAYEQKVNKIKEYPLKRAILHQENALYFRLPSNFDLLEKYLKTIWQVSKETPFSEMDFMFKNISEKIIFPERGSKMLDLKSNYKKcomplement (752. .1465)					
/gene="spr1057"					
complement (752. .1465)					
/gene="spr1057"					
/codon_start=1					
/transl_table=11					
/label=spr1057					
/product="Conserved hypothetical protein"					
/protein_id="AAK99861.1"					
/db_xref="GI:15458679"					
/translation="MFYKFLFLDLDHTLLDFFDAEDVALTQLIKERGVDIOAYKDYIVPMKALWKDLKLEKLVNFRSLRFLHFQGEKDSFLAQRQYFVLAQOGTSLSCAHLLDLSLIRDYDLVAENGLTATQIGRLAQSGLVFNQVFISEQLQKPAFLFYEKIQQIAGFSKSKTLMIGDLSLTADLQGGNAGIDITIWNPHHLENLQAOPTVEY					

Qy	241	TGAAGATTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATGTTAAATGA	300	Qy	1321	TGATAAGCATATATACTGTAACTGAGGCTCATAAAGCCCTTGTGTTGNAATAAAGGTCG	1380
Db	106710	TGAAGATTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATGTTAAATGA	106651	Db	105630	TGATAAGCATATATACTGTAACTGAGGCTCATAAAGCCCTTGTGTTGNAATAAAGGTCG	105571
Qy	301	GGTCAAGGGTGGATATGTTATCAAGCTAGATGGAATAATCTATGTTTACCTTAAAGGATGC	360	Qy	1381	TAAATCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATCGACTAA	1440
Db	106650	GGTCAAGGGTGGATATGTTATCAAGCTAGATGGAATAATCTATGTTTACCTTAAAGGATGC	106591	Db	105570	TAAATCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATCGACTAA	105511
Qy	361	TGCCACCGGATTAACGTCCTGATCAAAAGAGAGGAATCAATCGACAAAACAAGACATAG	420	Qy	1441	TAAAGAAAATTTGGTAGATGATTTATTTGGCATTCTTAGCACCATTACCCATCCGAGCG	1500
Db	106590	TGCCACCGGATTAACGTCCTGATCAAAAGAGAGGAATCAATCGACAAAACAAGACATAG	106531	Db	105510	TAAAGAAAATTTGGTAGATGATTTATTTGGCATTCTTAGCACCATTACCCATCCGAGCG	105451
Qy	421	TCAACATCGTAGAGTGGACCTCAAGAACGATGGTCTGCTTGGACGCTTGGCA	480	Qy	1501	ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTCAAGACGAAAGTTCGTAATTCCTCAAT	1560
Db	106530	TCAACATCGTAGAGTGGACCTCAAGAACGATGGTCTGCTTGGACGCTTGGCA	106471	Db	105450	ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTCAAGACGAAAGTTCGTAATTCCTCAAT	105391
Qy	481	AGGAGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	540	Qy	1561	AGCTGATTAAGTATACAACTGATGATGTTTACATTTTGTGATGACATGATATATCAGTCA	1620
Db	106470	AGGAGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC	106411	Db	105390	AGCTGATTAAGTATACAACTGATGATGTTTACATTTTGTGATGACATGATATATCAGTCA	105331
Qy	541	TGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACATTAATTCCTTAAGATGAGTT	600	Qy	1621	TGAAGAGATGATGATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG	1680
Db	106410	TGGTGATGCTTATATCGTTCCTCATGGAGATCATTAACATTAATTCCTTAAGATGAGTT	106351	Db	105330	TGAAGAGATGATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAG	105271
Qy	601	ATCAGCTAGCGATTTGGCTGCTGCAAGACGCTTCTATCTGCTGAGGAAATCTGTCAA	660	Qy	1681	CTTTCTTGATAAGGAAAAGTTGCAAGCTCAAGCCTTATCTAAGAAAAGGATTCCTACC	1740
Db	106350	ATCAGCTAGCGATTTGGCTGCTGCAAGACGCTTCTATCTGCTGAGGAAATCTGTCAA	106291	Db	105270	CTTTCTTGATAAGGAAAAGTTGCAAGCTCAAGCCTTATCTAAGAAAAGGATTCCTACC	105211
Qy	661	TTCAAGAACCTATCGCGCAACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC	720	Qy	1741	TCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGAGATAGTGCAGCAGCTATTTA	1800
Db	106290	TTCAAGAACCTATCGCGCAACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC	106231	Db	105210	TCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGAGATAGTGCAGCAGCTATTTA	105151
Qy	721	TGTAAGCAATCCAGAACTACAAATCTAACAAGCAACAAGCAACAACCTAACAGTCA	780	Qy	1801	CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA	1860
Db	106230	TGTAAGCAATCCAGAACTACAAATCTAACAAGCAACAAGCAACAACCTAACAGTCA	106171	Db	105150	CAATCGTGTGAAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA	105091
Qy	781	AGCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	840	Qy	1861	TACAGTTGAGGTTTAAAGCGGTAATTTGATTTATTCCTCATAGGATCATTAACATATAT	1920
Db	106170	AGCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	106111	Db	105090	TACAGTTGAGGTTTAAAGCGGTAATTTGATTTATTCCTCATAGGATCATTAACATATAT	105031
Qy	841	TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	900	Qy	1921	TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1980
Db	106110	TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	106051	Db	105030	TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	104971
Qy	901	AGCTAGAGGTTGCGAGTCCACACACGAGATCATTAACCACTTATCTCCCTTACTCTCAA	960	Qy	1981	TTTGTGTCGACGATTAAGTACTTACGTAGAACACCTCGACACCTCGACACCTCGAC	2040
Db	106050	AGCTAGAGGTTGCGAGTCCACACACGAGATCATTAACCACTTATCTCCCTTACTCTCAA	105991	Db	104970	TTTGTGTCGACGATTAAGTACTTACGTAGAACACCTCGACACCTCGACACCTCGAC	104911
Qy	961	GTCTGAATTTGGAGAACGATCGCTGATATTTCCCTTCTGTTTATCTCAACCATTTG	1020	Qy	2041	TGGATGGGCAATGCGAGTGGAGCATGTTAGGCAAGAACACACAGTGAAGATCCAA	2100
Db	105990	GTCTGAATTTGGAGAACGATCGCTGATATTTCCCTTCTGTTTATCTCAACCATTTG	105931	Db	104910	TGGATGGGCAATGCGAGTGGAGCATGTTAGGCAAGAACACACAGTGAAGATCCAA	104851
Qy	1021	GGTACAGATTTCAAGGCCAGAACCAACAGTCCACACACGACTCCGGAACCTAGTCCAG	1080	Qy	2101	TAAAGATTTCAAGCCGATGAGAGCGAGTGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	2160
Db	105930	GGTACAGATTTCAAGGCCAGAACCAACAGTCCACACACGACTCCGGAACCTAGTCCAG	105871	Db	104850	TAAAGATTTCAAGCCGATGAGAGCGAGTGTAGAGGAAACACCTGCTGAGCCAGAGTCCC	104791
Qy	1081	CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGGT	1140	Qy	2161	TCAAGTAGAGATGTAAGGAGTGTAGAGCCCACTCAAGAGAGCAGAAAGTTTGTGTCGAA	2220
Db	105870	CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGGT	105811	Db	104790	TCAAGTAGAGATGTAAGGAGTGTAGAGCCCACTCAAGAGAGCAGAAAGTTTGTGTCGAA	104731
Qy	1141	ACGAAAGTTGGGGAAGGATATGATTTGGAAGAAAGGGGATCTCTGTTTATGCTTTGTC	1200	Qy	2221	AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGTTTACGAATAA	2280
Db	105810	ACGAAAGTTGGGGAAGGATATGATTTGGAAGAAAGGGGATCTCTGTTTATGCTTTGTC	105751	Db	104730	AGTAAAGGATTTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGTTTACGAATAA	104671
Qy	1201	GAAAGATTTACCATCTGAAATCTTTAAATCTTTGAAAGCAAGTTATCAAAAACAAGAG	1260	Qy	2281	TTTGTGCTCTTCAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG	2340
Db	105750	GAAAGATTTACCATCTGAAATCTTTAAATCTTTGAAAGCAAGTTATCAAAAACAAGAG	105691	Db	104670	TTTGTGCTCTTCAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG	104611
Qy	1261	TGTTTACACATTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATG	1320	Qy	2341	GTGTTTAAAGGAGGATTAATCTTCTCATCTGTAAGTAAAGGAGGAGGAGGAGGAGG	2389
Db	105690	TGTTTACACATTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATG	105631	Db	104610	GTGTTTAAAGGAGGATTAATCTTCTCATCTGTAAGTAAAGGAGGAGGAGGAGGAGG	104562

QY	961	GTCTGAATTCGAAAGAAAGAAATCGCTCGTATTATTCCTCCCTTCGGTTATCGTTCAAAACCAATTG	1020
DB	6777	GTCTGAATTCGAAAGAAAGAAATCGCTCGTATTATTCCTCCCTTCGGTTATCGTTCAAAACCAATTG	6718
QY	1021	GGTACCAAGATTCAAGGCGCAGAACCAACCAAGTCCCAACCCGACTCGGAACTAGTCCAGG	1080
DB	6717	GGTACCAAGATTCAAGGCGCAGAACCAACCAAGTCCCAACCCGACTCGGAACTAGTCCAGG	6658
QY	1081	CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTTCAGCTGGT	1140
DB	6657	CCCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTTCAGCTGGT	6598
QY	1141	ACGAAAGTTGGGGAAGGATATGATTTCGAAGAAAGGCAATCTCTCGTTATGCTTTGTC	1200
DB	6597	ACGAAAGTTGGGGAAGGATATGATTTCGAAGAAAGGCAATCTCTCGTTATGCTTTGTC	6538
QY	1201	GAAAGATTTACCATCTGAACCTGTTTAAAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG	1260
DB	6537	GAAAGATTTACCATCTGAACCTGTTTAAAAATCTTTGAAAGCAAGTTATCAAAACAAGAGAG	6478
QY	1261	TGTTTTCACACTTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAAATTTTA	1320
DB	6477	TGTTTTCACACTTTAACTGCTAAAAAGAAAAATGTTGCTCCTCGTGACCAAGAAATTTTA	6418
QY	1321	TGATAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGTTGNAATTAAGGTCG	1380
DB	6417	TGATAAGCATATAATCTGTTTAACTGAGGCTCATAAAGCCTTGTTGNAATTAAGGTCG	6358
QY	1381	TAAATCTGATTTCCAAAGCCTTTAGACAATATTATTAGAACGCTTGAAATGATGAATCGACTAA	1440
DB	6357	TAAATCTGATTTCCAAAGCCTTTAGACAATATTATTAGAACGCTTGAAATGATGAATCGACTAA	6298
QY	1441	TAAAGAAAAATTTGGTAGATGATTTATTGGCAATTCCTAGCAACCAATTACCCTCCAGAGCG	1500
DB	6297	TAAAGAAAAATTTGGTAGATGATTTATTGGCAATTCCTAGCAACCAATTACCCTCCAGAGCG	6238
QY	1501	ACTTGGCAAAACCAATTCCTCAAAATGAGTATCTGAAGACGAAGTTCGTATTGCTCAAT	1560
DB	6237	ACTTGGCAAAACCAATTCCTCAAAATGAGTATCTGAAGACGAAGTTCGTATTGCTCAAT	6178
QY	1561	AGCTGATAAGTATACAACGTCAGATGTTTACATTTTGTATGAACATGATATAATCAGTGA	1620
DB	6177	AGCTGATAAGTATACAACGTCAGATGTTTACATTTTGTATGAACATGATATAATCAGTGA	6118
QY	1621	TGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTGCAATTTGGAAGAGATAG	1680
DB	6117	TGAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCACTGCAATTTGGAAGAGATAG	6058
QY	1681	CCTTCTCGATTAAGGAAAAAGTTCCAGCTCAAGCCTATACCTGCAATTTGGAAGAGATAG	1740
DB	6057	CCTTCTCGATTAAGGAAAAAGTTCCAGCTCAAGCCTATACCTGCAATTTGGAAGAGATAG	5998
QY	1741	TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGAGCTATTTA	1800
DB	5997	TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGAGCTATTTA	5938
QY	1801	CAATCTGTTGAAAGGGGAAAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA	1860
DB	5937	CAATCTGTTGAAAGGGGAAAAAGAAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA	5878
QY	1861	TACAGTTGAGTTTAAAAACGGTAAATTTGATATTTCCTCATAAGGATCATTACCATATAT	1920
DB	5877	TACAGTTGAGTTTAAAAACGGTAAATTTGATATTTCCTCATAAGGATCATTACCATATAT	5818
QY	1921	TAAATTTGCTTGGTTTGGATGATCACAATCAAAAGCTCCAAATGGCTATACCTTGGAGA	1980
DB	5817	TAAATTTGCTTGGTTTGGATGATCACAATCAAAAGCTCCAAATGGCTATACCTTGGAGA	5758
QY	1981	TTTGGTTGGGACGATTAAGTACTACGTAGAACACCTCGACCAAGCTCCCACTTCTTAATGA	2040
DB	5757	TTTGGTTGGGACGATTAAGTACTACGTAGAACACCTCGACCAAGCTCCCACTTCTTAATGA	5698
QY	2041	TGGATGGGCAATGCGCAGTGATGTGTTAGGCAAGAAAGACCCACAGTGAAGATCCAAA	2100

Db	5697	TGGATGGGCAATGCCAGTGGCATGTGTAGGCAAGAAGACCAAGTGAAGATCCAAA	5638
QY	2101	TAAGAATTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAAGTCCC	2160
Db	5637	TAAGAATTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCGAAGTCCC	5578
QY	2161	TCAAGTAGAGACTGAAGAAGTAGAGCCCACTCAAGAAGACGAAGTTTTCGTTGCGAA	2220
Db	5577	TCAAGTAGAGACTGAAGAAGTAGAGCCCACTCAAGAAGACGAAGTTTTCGTTGCGAA	5518
QY	2221	AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAAATCTAGCTGGTTTACGAAATAA	2280
Db	5517	AGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAAATCTAGCTGGTTTACGAAATAA	5458
QY	2281	TTTGACTCTTCAAAATATAGGATTAACATAGTATCATGGCAGAACGAAAAATTAATTGCG	2340
Db	5457	TTTGACTCTTCAAAATATAGGATTAACATAGTATCATGGCAGAACGAAAAATTAATTGCG	5398
QY	2341	GTTCGTTAAAAGGAAGTAATCCCTCATCTCTAGTAGGAGGAAAAATAAAC	2389
Db	5397	GTTCGTTAAAAGGAAGTAATCCCTCATCTCTAGTAGGAGGAAAAATAAAC	5349
RESULT 11			
AX571763/c			
LOCUS	AX571763	349980 bp	DNA
DEFINITION	Sequence 4982 from Patent WO02077021.	linear	PAT 29-NOV-2002
ACCESSION	AX571763		
VERSION	AX571763.1	GI:26003955	
KEYWORDS			
SOURCE	Streptococcus pneumoniae		
ORGANISM	Streptococcus pneumoniae		
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
AUTHORS	Streptococcus.		
TITLE	1		
JOURNAL	Masignani, V., Tettelin, H. and Fraser, C.		
FEATURES	Streptococcus pneumoniae proteins and nucleic acids		
source	Patent: WO 02077021-A 4982 03-OCT-2002;		
	Chiron Spa (IR); THE INSTITUTE FOR GENOMIC RESEARCH (US)		
	Location/Qualifiers		
	1..349980		
	/organism="Streptococcus pneumoniae"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:1313"		
	/note="seq 4979 too long: 2.162.598 bases-replaced by		
	following seq;-seq 4979: from 0.000.001 to 0.349.980-seq		
	4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001		
	to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq		
	4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001		
	to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq		
	4986: from 2.100.001 to 2.162.598"		
ORIGIN			
Query Match			
Best Local Similarity 97.9%; Score 2338; DB 6; Length 349980;			
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	TTCTTACGAGTTGGGACGTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA	60
Db	106950	TTCTTACGAGTTGGGACGTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA	106891
QY	61	TATAGATGGAACCAAGCGACGCAAAAACGGAGAAATTCGATCCCTGATGAGGTAGCAA	120
Db	106890	TATAGATGGAACCAAGCGACGCAAAAACGGAGAAATTCGATCCCTGATGAGGTAGCAA	120
QY	121	GGGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAAGCTATGTCAC	180
Db	106830	GGGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAAGCTATGTCAC	180
QY	181	TTACATCGGCACCACTATCAATTAATTAATGGTAAGGTTCCCTATGACCTATCATCAG	240
Db	106770	TTACATCGGCACCACTATCAATTAATTAATGGTAAGGTTCCCTATGACCTATCATCAG	106711

```

/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75283.1"
/db_xref="GI:14972654"
/translation="MKINKKVLGASVAVLALSVCSYELGRYQAGQDKKESNRVAYIDG
DOAGKAEHLTPDVRKREGINAGQIVIKITDQGVVTSBGHSHYHYNGKVPYDAIIE
ELLMKDPNYQLKSDIUNEIKGGVYKNGVYVYLKDAHADNIRTKKEIKROKQER
SHHNSRADNAVAARAGQRTYDDGYIPNASDIILEDGDAIVPHGDHVIKPNEL
SASLAAABAAVQRGSRSSSSSYNANPAQRLSENNHNLITVPTHQNGENISL
RIELYAKPLSHWSDRPEEPSPOPTPEPSPPOPPAPNPTDEKLKVEAVRKGVDGYV
RIILRYKSNHWSDRPEEPSPOPTPEPSPPOPPAPNPTDEKLKVEAVRKGVDGYV
FEENGVSRIYPAKDISAATAAGIDSKLAQBSLHKLTKTKTDLPSSDREFFNYKAYDL
LARHQDLNNGKQVDFEALDNLRLERLKVSSDKVLVEDILAFAPLIRPERLGP
NAQIYTDDEIQAVALAGKYTEDGYIFDPDILTSDEGDVVTMHTSHWLDPMYNLOY
EAEARAAQYAKAEKGLTPSDHQSGETEAKGAEAIYNRVKAAKVHPLDRMPNLSY
TEVKGSLIIPHYDHYNIKFEPFDEGLYAPKGYTILEDLLATLVKYIVHPNRPHS
DNFGFNASDHWQRNKGADNQTKEPSEKPOTEKPEETPREKPOSEKPEPKPT
REPESPESEPOVETEKVEKLEAEADLLGIQDPIIKSNKAEKTLGLKNNLLFGT
QDNNTIMAEAEKLLALKES"
complement(5346..7754)
/gene="Sp1175"
complement(5346..7754)
/gene="Sp1175"
/translation="MILSVCSYELGLYQARTVKNRVSVIDGKQATOKTENLTPDEV
IYNEKGINAQIVIKITDQGVVTSBGHSHYHYNGKVPYDAIIEELMLKDPNYKLKDED
SKREGINAGQIVIKITDQGVVTSBGHSHYHYNGKVPYDAIIEELMLKDPNYKLKDED
IYNEKGINAQIVIKITDQGVVTSBGHSHYHYNGKVPYDAIIEELMLKDPNYKLKDED
ALARSQRYTDDGYIPNASDIILEDGDAIVPHGDHSHYIIPNELNSBELAAAEALF
SGRGLNSLRVRRQNSDNTSNTVPSNFTNTSNTSNTSNTSNTSNTSNTSNTSNTSNTS
LKLQVLPUSQSHVEDGLVFPDPAQITSTARGAVPHGDHSHYIIPYQMSSELERIA
RILPLRYSNHWSDRPEEPSPOPTPEPSPPOPPAPNPTDEKLKVEAVRKGVDGYV
YYPEKGLISRYVAKDLPSSETVKNLESKLSKOSVSHTLTAKENVAPRQDFYKAY
NULTTAHKAFLFENKGRNSDQFQALDKLLERLINESKLEKLVDDLLAFAPLIRPERL
KENSQIYETVEDRVAQLADKYTSDGYIFDPSHDIISDEGDVVTMHTSHWLDPMYNLOY
LSDKQVAAQYTKGKGLIPSPDADKANPTGDSAAIYNRVKGKRIPLVRLPVMV
BHVTVKNGNLIIPKHQVHNTKPAFDHDTYKAPNGYTLDELFAFIKYVVEHPDERP
HSDNGWNASHEVLGKDHSDENPKNFKADEPVEETPAEPVPOVETEKVEAQLKEA
EVLLAKVTSSLSKANATETLAGLNLLTLQIMDNNSIMAEAEKLLALKSGNSPVSVK
EKIN"
complement(8049..9782)
/gene="Sp1176"
complement(8049..9782)
/gene="Sp1176"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAK75285.1"
/db_xref="GI:14972656"
/translation="MTEMLKGIASDGVAVAKAYLLVOPDLSPETITVEDTNAEEARL
DALQASQELSVIREKAVGTGCEAAQVDFDAHLMVADPEMISOIKETIRAKKNVAE
AGLKEVDMFTITIFEGEDNPNYQERAAADIRVTKVLANLLGKLLPNPASINEEVIV
ILADLTPSDTAQDKNFVAFVNIIGRTSHSAIMARTLEIAAVTGNNTITELVKQGD
ILAVNGITGEVINPTDEQAAEFKAAAGEAYAKQAEWALLKDAQVTADGKHFEELAN
IGTPKDVGNNGNEBAVGLYRTEFLYMDSQDFPTEDEQYAYKAVLGNMGKPVVVR
TMDIGDKELPYFDMPEHNPFLGFRALRISISPTEDGDMERTQIRALLRASVHGQLRI
MFMVALLKEPRAKAVDFEKKANLLAEGVAVADNIQVGMIEIIPAAAMADOPFAKEV
DFSIGTNDLIQTMAADRNEQVSLYQYNPISILRLINNVIKAAHAEGKWAGMGE
MAGDQQAVALLVGMGDFSMATSVLIRLSMKLLDTAKMEYANRALLTECSTMEEV
LELQKEYVNF"
complement(9788..10051)
/gene="Sp1177"
complement(9788..10051)
/gene="Sp1177"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein Hpr"
/protein_id="AAK75286.1"

```

gene

CDS

gene

CDS

gene

CDS

```

/db_xref="GI:14972657"
/translation="WASKDFHVVAETGIHARPATLLVOTASKFASDITLEYKGSVNL
KXINGVMSLGVGGQADVITISAGADADDAIAAISSETMEKEGLA"

```

ORIGIN

	Query Match	Best Local	Mismatches	Conservative	Score	DB 1	Length	Gaps	Indels
Qy	1	TTCCTACAGATTGGGACHTGTATCAAGCTAGAACCGTAAAGGAAAAATAATCGTGTTCCTA	60						
Db	7737	TTCCTACAGATTGGGACHTGTATCAAGCTAGAACCGTAAAGGAAAAATAATCGTGTTCCTA	7678						
Qy	61	TATAGATGGAACAAAGCGACGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA	120						
Db	7677	TATAGATGGAACAAAGCGACGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA	7618						
Qy	121	GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGTAACAGACCAAGGCTATGTCAC	180						
Db	7617	GGGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGTAACAGACCAAGGCTATGTCAC	7558						
Qy	181	TTCACATGGCGACCACTATCATTATTAACAATGGTAAGGTTCTTATGACGCTATCATCAG	240						
Db	7557	TTCACATGGCGACCACTATCATTATTAACAATGGTAAGGTTCTTATGACGCTATCATCAG	7498						
Qy	241	TGAAGAATTACTCATGAAAGATCCAAACCTATAGCTTAAAGATAGGATATTGTTAATGA	300						
Db	7497	TGAAGAATTACTCATGAAAGATCCAAACCTATAGCTTAAAGATAGGATATTGTTAATGA	7438						
Qy	301	GGTCAAGGTTGATATGTTTCAAGGTAGATCGAAATATCTATGTTTACCTTAAGGATGC	360						
Db	7437	GGTCAAGGTTGATATGTTTCAAGGTAGATCGAAATATCTATGTTTACCTTAAGGATGC	7378						
Qy	361	TGCCCCAGCGGATAACGTCCTGACAAAAGAGAAATCAATCGACAAAACAGAGCATAG	420						
Db	7377	TGCCCCAGCGGATAACGTCCTGACAAAAGAGAAATCAATCGACAAAACAGAGCATAG	7318						
Qy	421	TCAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTGTCCTTGGCAGGTTGCGCA	480						
Db	7317	TCAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTGTCCTTGGCAGGTTGCGCA	7258						
Qy	481	AGGACGCTATCTACAGATGATGGTTATCTTTAATGCTTCTGATATCATAGAGATAC	540						
Db	7257	AGGACGCTATCTACAGATGATGGTTATCTTTAATGCTTCTGATATCATAGAGATAC	7198						
Qy	541	TGCTGATGCTTATATCGTTCCTCATCGAGATCATTAACCAATTCCTAAGAAATGAGTT	600						
Db	7197	TGCTGATGCTTATATCGTTCCTCATCGAGATCATTAACCAATTCCTAAGAAATGAGTT	7138						
Qy	601	ATCAGCTAGCGAGTTGGCTGCTCAGAGCCTTCTTATCTGCTCGAGGAAATCTGTCAA	660						
Db	7137	ATCAGCTAGCGAGTTGGCTGCTCAGAGCCTTCTTATCTGCTCGAGGAAATCTGTCAA	7078						
Qy	661	TTCAGAACCTATCGCGCAAAAATAGCGATACACTTCAAGAACAACTGGGTACCTTC	720						
Db	7077	TTCAGAACCTATCGCGCAAAAATAGCGATACACTTCAAGAACAACTGGGTACCTTC	7018						
Qy	721	TGTAAGCAATCCAGGAACACTACAAATACTAACAAAGCAACAGCAACACTAACAGTCA	780						
Db	7017	TGTAAGCAATCCAGGAACACTACAAATACTAACAAAGCAACAGCAACACTAACAGTCA	6958						
Qy	781	AGCAAGTCAAAGTAATGACATTTGATGCTCTTTGAAACAGCTCTACAACTGCTTTGAG	840						
Db	6957	AGCAAGTCAAAGTAATGACATTTGATGCTCTTTGAAACAGCTCTACAACTGCTTTGAG	6998						
Qy	841	TCAACGATGATGAAATCTGATGCTTGTCTTGTTCAGCAAAATCACAAGTCGAAC	900						
Db	6897	TCAACGATGATGAAATCTGATGCTTGTCTTGTTCAGCAAAATCACAAGTCGAAC	6838						
Qy	901	AGCTAGAGGTTGTCAGTGGCCACACGAGATCATTAACCTTATCCCTTACTCTCAAT	960						
Db	6837	AGCTAGAGGTTGTCAGTGGCCACACGAGATCATTAACCTTATCCCTTACTCTCAAT	6778						



Db	4853	CAATCGTGTGAAGGGGAAACGAAATCCACTGGTTCGACTTCATATATGGTTGACCA	4912
Qy	1861	TACAGTTGAGGTTAAACCGTAATTTGATTAATTCCTCATAGGATCATACCAATAAT	1920
Db	4913	TACAGTTGAGGTTAAACCGTAATTTGATTAATTCCTCATAGGATCATACCAATAAT	4972
Qy	1921	TAAATTTGCTTGGTTTGAATGATCACATACAAAGCTCCCAATGGCTATACCTTGGAGA	1980
Db	4973	TAAATTTGCTTGGTTTGAATGATCACATACAAAGCTCCCAATGGCTATACCTTGGAGA	5032
Qy	1981	TTTGTTCGCGAGTAATAGTACTAGTAGAACCCTGACGAACGTCACATTCCTAATGA	2040
Db	5033	TTTGTTCGCGAGTAATAGTACTAGTAGAACCCTGACGAACGTCACATTCCTAATGA	5092
Qy	2041	TGGATGGGGAATGCCAGTAGCATGTGTAGGCAAGAACACACAGTGAAGATCCAAA	2100
Db	5093	TGGATGGGGAATGCCAGTAGCATGTGTAGGCAAGAACACACAGTGAAGATCCAAA	5152
Qy	2101	TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTCTGAGCCAGAGTCCC	2160
Db	5153	TAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTCTGAGCCAGAGTCCC	5212
Qy	2161	TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAACGAGAAGTTTGTCTGGAA	2220
Db	5213	TCAAGTAGAGACTGAAAGAGTAGAAGCCCAACTCAAAGAACGAGAAGTTTGTCTGGAA	5272
Qy	2221	AGTAACCGATCTAGTCTGAAGCGAATGCAACGAACTCTAGCTGGTTACGAATAA	2280
Db	5273	AGTAACCGATCTAGTCTGAAGCGAATGCAACGAACTCTAGCTGGTTACGAATAA	5332
Qy	2281	TTTGACTCTTCAAAATTTAGTATAACAATAGTATCATGCGAAGCAAGAAATTTACTTGC	2340
Db	5333	TTTGACTCTTCAAAATTTAGTATAACAATAGTATCATGCGAAGCAAGAAATTTACTTGC	5392
Qy	2341	GTGCTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGCAAGAAATAAAC	2389
Db	5393	GTGCTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGCAAGAAATAAAC	5441
RESULT 10			
AE007418/c			
LOCUS	AE007418	10256 bp	DNA linear BCT 31-AUG-2001
DEFINITION	Streptococcus pneumoniae TIGR4 section 101 of 194 of the complete genome.		
ACCESSION	AE007418	AE005672	
VERSION	AE007418.1	GI:14972649	
KEYWORDS	Streptococcus pneumoniae TIGR4		
SOURCE	Streptococcus pneumoniae TIGR4		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1 (bases 1 to 10256)		
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Unayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.		
	Complete genome sequence of a virulent isolate of Streptococcus pneumoniae		
JOURNAL	Science	293 (5529), 498-506	(2001)
MEDLINE	21357209		
PUBMED	11463916		
REFERENCE	2 (bases 1 to 10256)		
AUTHORS	Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Unayam, L.A., White, O., Salzberg, S.L., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.		

TITLE	JOURNAL	Submitted (29-JUN-2001)	The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers		
source	1..10256		
	/organism="Streptococcus pneumoniae TIGR4"		
	/mol_type="genomic DNA"		
	/strain="TIGR4"		
gene	/db_xref="taxon:170187"		
	/complement(100..702)		
CDS	/gene="SP1170"		
	/complement(100..702)		
	/gene="SP1170"		
	/notes="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK75279.1"		
	/db_xref="GI:14972650"		
	/translation="MSEVDFNEAVNYEFTSDTCQLANSIYQSLEKFFDKNFSGDLIF		
	TKPSLVKGGDYIGRRDQVDNLRLVIGNIFPNLNRKYSLNMRNGCMGDFPHDFF		
	DIYLDHVAKYAVQKNNIKEYVPLKRAILHQENALYERFESNFDLLEKNYLTQW		
	VSEKTFSEMDFMNPKNISEKLIIFERGSMLNDLKSNNK"		
	/complement(760..1473)		
gene	/gene="SP1171"		
	/complement(760..1473)		
CDS	/gene="SP1171"		
	/notes="identified by match to PFAM protein family HMM		
	PF00702"		
	/codon_start=1		
	/transl_table=11		
	/product="hydrolase, haloacid dehalogenase-like family"		
	/protein_id="AAK75280.1"		
	/db_xref="GI:14972651"		
	/translation="MFYKELLDLHLLDFAAEDVALTQLLKESGVADIQAKDYV		
	VPNKALWDLEKKISKQELVNRFRSLFAHQKQSGFLAQRFYLAQKQQLS		
	GAHLLDLSLIERDNLNATNGITATQGLAQSGLAPEYVFISEQLTQKQDALF		
	YXIQQIAGFSKSKTLMIGDSLTDIQGNNAGIDTIWNPHEHNTQAQPTVEY		
	SYQQLDCLDKNLEKITP"		
	/complement(1724..1843)		
gene	/gene="SP1172"		
	/complement(1724..1843)		
CDS	/gene="SP1172"		
	/notes="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK75281.1"		
	/db_xref="GI:14972652"		
	/translation="MLYAVPFYFNRSETIVFLNCESIKTDCGAILALETFKN"		
	/complement(1933..2307)		
gene	/gene="SP1173"		
	/complement(1933..2307)		
CDS	/gene="SP1173"		
	/notes="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved hypothetical protein"		
	/protein_id="AAK75282.1"		
	/db_xref="GI:14972653"		
	/translation="MESIGLVISHSKHIAEGVWELISKVAKDVPITYVRGEGGIG		
	TSFEQVDRVSVENPADTLAFLDGLSAKNLKNVTFDSKSIINRPIVEGAYNAA		
	LIQAGAEISVIQTQLAELEINK"		
	/complement(2729..5188)		
gene	/gene="SP1174"		
	/complement(2729..5188)		
CDS	/gene="SP1174"		
	/notes="similar to GP:4249624; identified by sequence		
	similarity; putative"		
	/codon_start=1		

PF	30-OCT-1997 JP 1998520718
PR	31-OCT-1996 US 60/029960
PI	CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH.
PI	MICHAEL FANNON, BRIAN A DOUGHERTY
PC	C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC	C12N1/21,
PC	C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40	
CC	Strandedness: Double;
CC	Topology: Linear;
PH Key	Location/Qualifiers
FT source	1..8195
FT	Location/Qualifiers
FEATURES	/organism='Unidentified'.
source	1..8195
	/organism='unidentified'
	/mol_type='genomic DNA'
	/db_xref='taxon:32644'
ORIGIN	
Query Match	97.9%; Score 2338; DB 6; Length 8195;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2388;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1	TTCTTACGACTTCGGACTGTATCAAGCTGAACCGTTTAAGGAAAATAAATCGTGTTCCCTA 60
Db	
3053	TTCCTTAGAGTTGGAGCTGTATCAAGCTGAACCGTTTAAGGAAAATAAATCGTGTTCCCTA 3112
QY	61 TATAGATGGAAAAACAACGCGACGCAGAAACCGAGAAATTTCACCTCCTGATGAGGTTAGCAA 120
Db	
3113	TATAGATGGAAAAACAACGCGACGCAGAAACCGAGAAATTTCACCTCCTGATGAGGTTAGCAA 3172
QY	121 GCGTGAAGGAATCAATCTGAGCMAATTCGTCATCAGATAACAGACCAGGCTATGTCAC 180
Db	
3173	GCGTGAAGGAATCAATCTGAGCMAATTCGTCATCAGATAACAGACCAGGCTATGTCAC 3232
QY	181 TTCACATGGCGACCACTATCATTTATTAACAATGGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db	
3233	TTCACATGGCGACCACTATCATTTATTAACAATGGTAAGGTTCCCTTATGACGCTATCATCAG 3292
QY	241 TGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAAGATGAGGATATTTGTTAATGA 300
Db	
3293	TGAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAAGATGAGGATATTTGTTAATGA 3352
QY	301 GGTCAAAGGTTGGATATGTTTCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 360
Db	
3353	GGTCAAGGGTGGATATGTTTCAAGGTAGATGGAATAATACTATGTTTACCTTAAGGATGC 3412
QY	361 TGCCCCACGCGATAAACCGTCCGTCACAAAGAGGAAATCAATCGACAAAACAAGAGCATAG 420
Db	
3413	TGCCCACGCGATAAACCGTCCGTCACAAAGAGGAAATCAATCGACAAAACAAGAGCATAG 3472
QY	421 TCACATCTGGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTGTCCTTGGCAGCGTTCGCA 480
Db	
3473	TCACATCTGGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTGTCCTTGGCAGCGTTCGCA 3532
QY	481 AGGACGCTATACTACAGATCATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
Db	
3533	AGGACGCTATACTACAGATCATGTTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 3592
QY	541 TGGTGATGCTTATATCGTTCTCTCATAGGATCATTAACATTAACATTCCTTAAGAAATGAGTT 600
Db	
3593	TGGTGATGCTTATATCGTTCTCTCATAGGATCATTAACATTCCTTAAGAAATGAGTT 3652
QY	601 ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCCTATCTGTCGAGGAATCTGTCAA 660
Db	
3653	ATCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCCTATCTGTCGAGGAATCTGTCAA 3712
QY	661 TTCAAGAACCTTATCGCCGCACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 720
Db	
3713	TTCAAGAACCTTATCGCCGCACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTC 3772



Db 3533 AGGACGCTACTACAGATGATGGTTATATCTTTAAATGCTTCTGATATCATAGAGATAC 3592  
Qy 541 TGGTGAATCTTATATCGTTCTCATGGAGATCATTAACATTTACATTTCTTAAGAAATGAGTT 600  
Db 3593 TGGTGAATCTTATATCGTTCTCATGGAGATCATTAACATTTACATTTCTTAAGAAATGAGTT 3652  
Qy 601 ATCAGCTAGCGAGTTGGCTGTCGAGAGCCCTTCTATCTGGTCGAGAAATCTGTCAAA 660  
Db 3653 ATCAGCTAGCGAGTTGGCTGTCGAGAGCCCTTCTATCTGGTCGAGAAATCTGTCAAA 3712  
Qy 661 TTCAAGAACCTTATCGCGACAAAATAGCGATAACACATTTCAAGAACAACTGGGTACCTTC 720  
Db 3713 TTCAAGAACCTTATCGCGACAAAATAGCGATAACACATTTCAAGAACAACTGGGTACCTTC 3772  
Qy 721 TGTAAAGCAATCCAGGAATCAAAATCTAACAAGCAAGCAACAGCAAGCAAGCAAGCAAGCA 780  
Db 3773 TGTAAAGCAATCCAGGAATCAAAATCTAACAAGCAAGCAACAGCAAGCAAGCAAGCAAGCA 3832  
Qy 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAATCGCTTTTGA 840  
Db 3833 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAATCGCTTTTGA 3892  
Qy 841 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 900  
Db 3893 TCAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCAAAATCAAGTCGAAC 3952  
Qy 901 AGCTAGAGGTGTTCCAGTGCACACGAGATCATTTACATTTCTTCTTCTTCTTCTTCTT 960  
Db 3953 AGCTAGAGGTGTTCCAGTGCACACGAGATCATTTACATTTCTTCTTCTTCTTCTTCTT 4012  
Qy 961 GTCTGAATTTGAAGCAAGTATGCTGATTTATTTCCCTTCTGTTTATCGTTTCAAACTT 1020  
Db 4013 GTCTGAATTTGAAGCAAGTATGCTGATTTATTTCCCTTCTGTTTATCGTTTCAAACTT 4072  
Qy 1021 GGTACCAAGTCAAGGCGAGCAACCAAGTCCAAACCGACTCCGAACTTCTGAGTCCAGG 1080  
Db 4073 GGTACCAAGTCAAGGCGAGCAACCAAGTCCAAACCGACTCCGAACTTCTGAGTCCAGG 4132  
Qy 1081 CCGCAACCTGCACCAATCTTAAATAGATCTCAATTTCTTCTTCTTCTTCTTCTTCTTCT 1140  
Db 4133 CCGCAACCTGCACCAATCTTAAATAGATCTCAATTTCTTCTTCTTCTTCTTCTTCTTCT 4192  
Qy 1141 ACAGAAAGTTGGGAAGGATATGATTTCAAGAAAGGGGATCTCTGTTATGCTTTTGC 1200  
Db 4193 ACAGAAAGTTGGGAAGGATATGATTTCAAGAAAGGGGATCTCTGTTATGCTTTTGC 4252  
Qy 1201 GAAAGATTTACCATCTGAACTGTAAATCTTTGAAAGCAAGTTATCAAAACAAAGAG 1260  
Db 4253 GAAAGATTTACCATCTGAACTGTAAATCTTTGAAAGCAAGTTATCAAAACAAAGAG 4312  
Qy 1261 TGTTCACACATTTAATCTGCTTAAAGAAAGTGTCTCTCTGTCGACCAAGATTTTA 1320  
Db 4313 TGTTCACACATTTAATCTGCTTAAAGAAAGTGTCTCTCTGTCGACCAAGATTTTA 4372  
Qy 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGAAATTAAGGGTGC 1380  
Db 4373 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGAAATTAAGGGTGC 4432  
Qy 1381 TAAATCTGATTTCCAGCTTAGACAAATTTAGACGCTTGAATGATGATGATGATGATGATG 1440  
Db 4433 TAAATCTGATTTCCAGCTTAGACAAATTTAGACGCTTGAATGATGATGATGATGATGATG 4492  
Qy 1441 TAAAGAAATTTGGTATGATGATTTATGCGATCTTACGACCAATTTACCCATCCAGAGC 1500  
Db 4493 TAAAGAAATTTGGTATGATGATTTATGCGATCTTACGACCAATTTACCCATCCAGAGC 4552  
Qy 1501 ACTTGGCAACCAATTTCTCAATTTAGTATGATGATGATGATGATGATGATGATGATGAT 1560  
Db 4553 ACTTGGCAACCAATTTCTCAATTTAGTATGATGATGATGATGATGATGATGATGATGAT 4612  
Qy 1561 AGCTGATAGTATCAACGTCAGATGGTTACATTTTGTGATGATGATGATGATGATGATGATG 1620  
Db 4613 AGCTGATAGTATCAACGTCAGATGGTTACATTTTGTGATGATGATGATGATGATGATGATG 4672

Qy 1621 TGAAGGAGATGCATATGTAAAGCCTCATATGGGCCATAGTCACTGGATTCGAAAAGATAG 1680  
Db 4673 TGAAGGAGATGCATATGTAAAGCCTCATATGGGCCATAGTCACTGGATTCGAAAAGATAG 4732  
Qy 1681 CCTTCTTGATAAGGAAAAGTTGCAGCTCAAGCCTATATAAAGAAAAGATATCTCTACC 1740  
Db 4733 CCTTCTTGATAAGGAAAAGTTGCAGCTCAAGCCTATATAAAGAAAAGATATCTCTACC 4792  
Qy 1741 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGACGATTTTA 1800  
Db 4793 TCCATCTCCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGACGATTTTA 4852  
Qy 1801 CAATCGTGTAAAGGGGAAAAGCGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 1860  
Db 4853 CAATCGTGTAAAGGGGAAAAGCGAATTCACCTCGTTCGACTTCCATATATGTTGAGCA 4912  
Qy 1861 TACAGTTGAGGTTTAAAAAACGGTAATTTGATTTATTTCTCATAAGGATCATTTACCATATAT 1920  
Db 4913 TACAGTTGAGGTTTAAAAAACGGTAATTTGATTTATTTCTCATAAGGATCATTTACCATATAT 4972  
Qy 1921 TAAATTTGCTTGGTTTGATGATCACATACAAAGCTTCCAAATGGCTATACCTTTGGAAGA 1980  
Db 4973 TAAATTTGCTTGGTTTGATGATCACATACAAAGCTTCCAAATGGCTATACCTTTGGAAGA 5032  
Qy 1981 TTTGTTTGGCAGCATTAAGTACTAGCTAGAACACCCCTGACGAAGCTCCACATCTTAATGA 2040  
Db 5033 TTTGTTTGGCAGCATTAAGTACTAGCTAGAACACCCCTGACGAAGCTCCACATCTTAATGA 5092  
Qy 2041 TGGATGGGGCAATGCCAGTGAAGTGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100  
Db 5093 TGGATGGGGCAATGCCAGTGAAGTGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAAA 5152  
Qy 2101 TAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCGCAAGTCCC 2160  
Db 5153 TAAGAACTTCAAAGCGGATGAAGCGCAGTAGAGGAAACACCTGCTGAGCGCAAGTCCC 5212  
Qy 2161 TCAAGTAGACACTCAAAAGTGAAGCCCAACTCAAGAGAGCAGAGTGTGTTGTTGCGAA 2220  
Db 5213 TCAAGTAGACACTCAAAAGTGAAGCCCAACTCAAGAGAGCAGAGTGTGTTGTTGCGAA 5272  
Qy 2221 AGTAAACGGATTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280  
Db 5273 AGTAAACGGATTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAA 5332  
Qy 2281 TTTGACTCTTCAAATTTAGATTAACAATAGTATCATGCGAGAGCAGAAAATTTACTTGC 2340  
Db 5333 TTTGACTCTTCAAATTTAGATTAACAATAGTATCATGCGAGAGCAGAAAATTTACTTGC 5392  
Qy 2341 GTTCTTAAAGGAAGTAATCCTTCATCTGTAAGTAAAGGAAAATAAAC 2389  
Db 5393 GTTCTTAAAGGAAGTAATCCTTCATCTGTAAGTAAAGGAAAATAAAC 5441

## RESULT 9

BD003774  
LOCUS BD003774 8195 bp DNA linear PAT 31-JAN-2002  
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.  
ACCESSION BD003774  
VERSION BD003774.1 GI:18631735  
KEYWORDS JP 2001501833-A/94.  
SOURCE unidentified  
ORGANISM Streptococcus pneumoniae  
REFERENCE 1 (bases 1 to 8195)  
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C., Fannon, M., and Dougherty, B.A.  
TITLE Polynucleotide of Streptococcus pneumoniae and sequence  
JOURNAL Patent: JP 2001501833-A 94 13-FEB-2001;  
HUMAN GENOME SCIENCES INC  
COMMENT OS Unidentified  
PN JP 2001501833-A/94  
PD 13-FEB-2001

Db 1098 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTCTTTGTTAGTCAGCTGGT 1157  
Qy 1141 ACGAAAGTTGGGAAGGATATGTTATCGAAGAAAGGCGCATCTCTCGTTATGCTTTTCG 1200  
Db 1158 ACGAAAGTTGGGAAGGATATGTTATCGAAGAAAGGCGCATCTCTCGTTATGCTTTTCG 1217  
Qy 1201 GAAAGATTTTACCATCTCGAACTCTTAAATCTTTGAAAGCAAGTTATCAAAAACAAGAG 1260  
Db 1218 GAAAGATTTTACCATCTCGAACTCTTAAATCTTTGAAAGCAAGTTATCAAAAACAAGAG 1277  
Qy 1261 TGTTTCACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGATTTTA 1320  
Db 1278 TGTTTCACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGATTTTA 1337  
Qy 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGCTG 1380  
Db 1338 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGCTG 1397  
Qy 1381 TAATCTGATTTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440  
Db 1398 TAATCTGATTTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1457  
Qy 1441 TAAAGAAAATTTGCTAGATGATTTATTGGCATCTCTAGCACCAATTAACCATCCAGAGCG 1500  
Db 1458 TAAAGAAAATTTGCTAGATGATTTATTGGCATCTCTAGCACCAATTAACCATCCAGAGCG 1517  
Qy 1501 ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGTTGCTATTGCTCAATT 1560  
Db 1518 ACTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAGTTGCTATTGCTCAATT 1577  
Qy 1561 AGCTGATAAGTATACAACTGATGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA 1620  
Db 1578 AGCTGATAAGTATACAACTGATGATGTTTACATTTTGTGATGAACATGATATAATCAGTGA 1637  
Qy 1621 TGAAGGAGATGATATGTTAAAGCCTCATATGAGGCTATGCTAGTGGAAAGATAG 1680  
Db 1638 TGAAGGAGATGATATGTTAAAGCCTCATATGAGGCTATGCTAGTGGAAAGATAG 1697  
Qy 1681 CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTATATCTAAAGAAAAAGGTTATCTTACC 1740  
Db 1698 CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCTATATCTAAAGAAAAAGGTTATCTTACC 1757  
Qy 1741 TCCATCTCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1800  
Db 1758 TCCATCTCAGACGCGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTA 1817  
Qy 1801 CAATCGTGTCAAGGGGAAAAAGCAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 1860  
Db 1818 CAATCGTGTCAAGGGGAAAAAGCAATTCACCTCGTTGCACTTCCATATATGTTGAGCA 1877  
Qy 1861 TACAGTTGAGTTTAAAGCGGTAATTTGATTTTCTCTATAAGGATCATTAACCATATAT 1920  
Db 1878 TACAGTTGAGTTTAAAGCGGTAATTTGATTTTCTCTATAAGGATCATTAACCATATAT 1937  
Qy 1921 TAAATTTGCTTGGTTTGGATGATCACATPACAAGCTCCAAATGGCTATACCTTGGGAAGA 1980  
Db 1938 TAAATTTGCTTGGTTTGGATGATCACATPACAAGCTCCAAATGGCTATACCTTGGGAAGA 1997  
Qy 1981 TTTGTTTGGAGATTAAGTACTAGTAGAACACCCCTGACGACGCTCCCATTTCTAATGA 2040  
Db 1998 TTTGTTTGGAGATTAAGTACTAGTAGAACACCCCTGACGACGCTCCCATTTCTAATGA 2057  
Qy 2041 TGGATGGGCAATGCGAGTGAGCTGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA 2100  
Db 2058 TGGATGGGCAATGCGAGTGAGCTGTTTAGCAAGAAAGACCAAGTGAAGATCCAAA 2117  
Qy 2101 TAAGAACTTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
Db 2118 TAAGAACTTTCAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2177  
Qy 2161 TCAAGTAGAGCTGAAGAAAGTAGAGCCCAACTCAAGAGCAGAGTTTTCGTTCCGAA 2220  
Db 2178 TCAAGTAGAGCTGAAGAAAGTAGAGCCCAACTCAAGAGCAGAGTTTTCGTTCCGAA 2237

Qy 2221 AGTAACGGATTTCTAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2280  
Db 2238 AGTAACGGATTTCTAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAA 2297  
Qy 2281 TTTGACTCTTCAAATTTATGGATAACAATAGTATATCGGCAGAGCAGAAAAATTTACTTGC 2340  
Db 2298 TTTGACTCTTCAAATTTATGGATAACAATAGTATATCGGCAGAGCAGAAAAATTTACTTGC 2357  
Qy 2341 GTTCTTAAAGGAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389  
Db 2358 GTTCTTAAAGGAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 2406

RESULT 8  
AR218862  
LOCUS AR218862  
DEFINITION Sequence 94 from patent  
ACCESSION AR218862  
VERSION AR218862.1 GI:23319796  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 8195)  
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,  
Fannon, M.R., and Dougherty, B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: US 6420135-A 94 16-JUL-2002;  
FEATURES  
Location/Qualifiers  
1..8195  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 97.9%; Score 2338; DB 6; Length 8195;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 TTTCTACGAGTTGGGACTGTTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 60  
Db 3053 TTTCTACGAGTTGGGACTGTTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTA 3112  
Qy 61 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCGATGAGGTTAGCAA 120  
Db 3113 TATAGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCGATGAGGTTAGCAA 3172  
Qy 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 180  
Db 3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAC 3232  
Qy 181 TTTCAATGCGGACCACTATCATTTACAAATGTAAGGTTTCTTATGAGCTATCATCAG 240  
Db 3233 TTTCAATGCGGACCACTATCATTTACAAATGTAAGGTTTCTTATGAGCTATCATCAG 3292  
Qy 241 TGAAGAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTCTTAATGA 300  
Db 3293 TGAAGAATTTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTTCTTAATGA 3352  
Qy 301 GGTCAAGGGTGGATATGTTTATCAAGTAGATGAAAAATACTATGTTTACTTAAGGATGC 360  
Db 3353 GGTCAAGGGTGGATATGTTTATCAAGTAGATGAAAAATACTATGTTTACTTAAGGATGC 3412  
Qy 361 TGCCCAACCGGATAAGCTCGTCAAAAAGAGGAATCAATCCAAAAACAGAGCATAG 420  
Db 3413 TGCCCAACCGGATAAGCTCGTCAAAAAGAGGAATCAATCCAAAAACAGAGCATAG 3472  
Qy 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACCATGGTGTCTGTTGCCCTGGCAGCTTCGCA 480  
Db 3473 TCAACATCGTGAAGTGGAACTCCAAAGAAACCATGGTGTCTGTTGCCCTGGCAGCTTCGCA 3532  
Qy 481 AGGACGCTATACATCAGATGATGGTTATCTTTTAAATGCTTCTGATATCATAGAGATAC 540

```
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACAGCTATTTA 1800
|
|
|
Db 1890 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACAGCTATTTA 1949
|
|
|
QY 1801 CAATCGTGTCAAAAGGGGAAAAAGCAATTCACATCGTTCGACTTCCATATATGTTGAGCA 1860
|
|
|
Db 1950 CAATCGTGTCAAAAGGGGAAAAAGCAATTCACATCGTTCGACTTCCATATATGTTGAGCA 2009
|
|
|
QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATATATTCCTCATAAGATCAATTAACATATAT 1920
|
|
|
Db 2010 TACAGTTGAGGTTAAAAACGGTAATTTGATATATTCCTCATAAGATCAATTAACATATAT 2069
|
|
|
QY 1921 TAAATTTGCTTGGTTTTCATGATCACACATACAAGCTCCAAATGGCTATACCTTGAAGA 1980
|
|
|
Db 2070 TAAATTTGCTTGGTTTTCATGATCACACATACAAGCTCCAAATGGCTATACCTTGAAGA 2129
|
|
|
QY 1981 TTTGTTTGGACGATTAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATTTCTAATGA 2040
|
|
|
Db 2130 TTTGTTTGGACGATTAAGTACTACGTAGAACACCCCTGACGAACGCTCCACATTTCTAATGA 2189
|
|
|
QY 2041 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2100
|
|
|
Db 2190 TGGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAA 2249
|
|
|
QY 2101 TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2160
|
|
|
Db 2250 TAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 2309
|
|
|
QY 2161 TCAAGTAGAGACTGAAAAGTGAAGGCCCAACTCAAAAGAGCAGAAAGTTTGTCTTGGCAA 2220
|
|
|
Db 2310 TCAAGTAGAGACTGAAAAGTGAAGGCCCAACTCAAAAGAGCAGAAAGTTTGTCTTGGCAA 2369
|
|
|
QY 2221 AGTAACGGATCTAGTCTGAAGCCATGCAAGAACTCTAGCTGGTTTACGAATATA 2280
|
|
|
Db 2370 AGTAACGGATCTAGTCTGAAGCCATGCAAGAACTCTAGCTGGTTTACGAATATA 2429
|
|
|
QY 2281 TTTGACTCTTCAAAATTTGGATAACAATAGTATCATGCGAAGCAGAAATAATTTACTTGC 2340
|
|
|
Db 2430 TTTGACTCTTCAAAATTTGGATAACAATAGTATCATGCGAAGCAGAAATAATTTACTTGC 2489
|
|
|
QY 2341 GTTGTTAAAGGAAGTAACTCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
|
|
|
Db 2490 GTTGTTAAAGGAAGTAACTCTTCATCTGTAAGTAAGGAAAAATAAAC 2538
|
|
|
RESULT 7
AX569139 2406 bp DNA linear PAT 29-NOV-2002
LOCUS Sequence 2347 from Patent WO02077021.
DEFINITION AX569139
ACCESSION AX569139
VERSION AX569139.1 GI:26002636
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1
AUTHORS Masignani, V., Tettelin, H. and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 2347 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source 1..2406
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"
ORIGIN
Query Match 97.9%; Score 2338; DB 6; Length 2406;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTTCCTA 60
```

```
Db 18 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTTTCCTA 77
|
|
|
QY 61 TATAGATGGAACAAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
|
|
|
Db 78 TATAGATGGAACAAACAGCGACGCAAAAAACGGAGAAATTTGACTCCTGATGAGGTTAGCAA 137
|
|
|
QY 121 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180
|
|
|
Db 138 GCGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 197
|
|
|
QY 181 TTCAATATGCGGACCACTATCATTTATCAATGTAAGGTTTCTTATCAGCTATCATCAG 240
|
|
|
Db 198 TTCAATATGCGGACCACTATCATTTATCAATGTAAGGTTTCTTATCAGCTATCATCAG 257
|
|
|
QY 241 TGAAGATTTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTTGTTAATGA 300
|
|
|
Db 258 TGAAGATTTACTCATGAAGATCCAACTATAAGCTTAAAGATGAGGATATTTGTTAATGA 317
|
|
|
QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC 360
|
|
|
Db 318 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGC 377
|
|
|
QY 361 TGCCCAACGCGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 420
|
|
|
Db 378 TGCCCAACGCGGATAACGTCGCTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAG 437
|
|
|
QY 421 TCAACATCTGTAAGGTGGAACTCAAGAAACGATGGTGTGCTTGCCTTGCGCATGTTGCGCA 480
|
|
|
Db 438 TCAACATCTGTAAGGTGGAACTCAAGAAACGATGGTGTGCTTGCCTTGCGCATGTTGCGCA 497
|
|
|
QY 481 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 540
|
|
|
Db 498 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 557
|
|
|
QY 541 TGTGATGCTTATATCGTTCCTCATGAGATCATACCATTTACATTCCTTAAGATGAGTT 600
|
|
|
Db 558 TGTGATGCTTATATCGTTCCTCATGAGATCATACCATTTACATTCCTTAAGATGAGTT 617
|
|
|
QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGNAGCCTTCTATCTGCTCGAGGAATCTGTCAA 660
|
|
|
Db 618 ATCAGCTAGCGAGTTGGCTGCTGAGNAGCCTTCTATCTGCTCGAGGAATCTGTCAA 677
|
|
|
QY 661 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720
|
|
|
Db 678 TTCAAGAACCTATCGCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 737
|
|
|
QY 721 TGTAAGCAATCCAGGAACCTACAAATACTAAACAAGCAACAAGCAACACTAACAGTCA 780
|
|
|
Db 738 TGTAAGCAATCCAGGAACCTACAAATACTAAACAAGCAACAAGCAACACTAACAGTCA 797
|
|
|
QY 781 AGCAAGTCAAGTAATGACATTCATGATGCTCTTGAACAGCTCTACAACTGCTTTGAG 840
|
|
|
Db 798 AGCAAGTCAAGTAATGACATTCATGATGCTCTTGAACAGCTCTACAACTGCTTTGAG 857
|
|
|
QY 841 TCAACGACATGTAGAATCTGATGCTTGTCTTTGATCCAGCACAATCAAGTTCGAAC 900
|
|
|
Db 858 TCAACGACATGTAGAATCTGATGCTTGTCTTTGATCCAGCACAATCAAGTTCGAAC 917
|
|
|
QY 901 AGCTAGAGGTTGTGAGTGCGCACGAGAGATCAATACCATTCATCCCTTACTCTCAAT 960
|
|
|
Db 918 AGCTAGAGGTTGTGAGTGCGCACGAGAGATCAATACCATTCATCCCTTACTCTCAAT 977
|
|
|
QY 961 GTCTGANTTGGAAAGCAAGATCGCTGTATTTATTCCTTCTGTTATCGTTTCAACCACTTG 1020
|
|
|
Db 978 GTCTGANTTGGAAAGCAAGATCGCTGTATTTATTCCTTCTGTTATCGTTTCAACCACTTG 1037
|
|
|
QY 1021 GGTACCAAGATTTCAAGGCGAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1080
|
|
|
Db 1038 GGTACCAAGATTTCAAGGCGAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGG 1097
|
|
|
QY 1081 CCGCAACCTGCGACCAAAATCTTAAATAGACTCAAATCTTCTTTGTTAGTCAAGCTGGT 1140
|
|
|
```

CDS	<div>/gene="phtA" 91..2541 /gene="phtA" /note="phtA" /codon_start=1 /transl_table=11 /product="pneumococcal histidine triad A protein" /protein_id="AAK19155.1" /db_xref="GI:13345013" /translation="MKINKYLVGSAAALLVCSVELGLYQARTVKENRVSVIDGK QATKENTLPDEVSREGINAEQIVIKITDQYVTSBGHDHYNGKVPYDALISBE LLMKDPNYLKDEDIVNEVKGQYVIKVDGKYVYLKAAAHADNVRTKEINRQKQHS QHREGGTPNDGADVALARGRYTDDGYIFNASDIIEDPTGDAYIVPHDGHYIIPKN ELASASELAEEAFPLSGRGNLSRTRYRQNSNTSRNWPVSNPNTNTNTNSNNS TNSQASOASQNDIDSLKQLVPLPSORHVESDGLVDPDAOITSTARGVAVPHGDVHF IPYSOMSELEERLARIIPLYRSNHWVPSRPEQSPQPTPEPSQPPAPNLKIDSN SSVGLVRKVGQYFEEEGISRYVFAKDLSPSEVNEKLSKQESVSHLTAKKE NVAPDQBYDGYNLLTEAHKLNKGRNSDFQALDKLERLNDENSTKEKLVDLL LAFLAPI THERLGPNSQI EYTEDVRIQAQADKYTSDGYIFDSDHDIISDSGDVAV TFMHGSHWIGKDSLSDKEKVAQAQYKTEKGLPPSPDADVKANPTGDSAAAIYNRVK GEKRIPLVRLPYMVEHTVEVKNGLIIPKDHVHLKFAWFDHNYKAENGYTLDELFI ATIKYVVEHPDRPHSDNGNASEHLVGGKHSDEPNKFKADBEPEVETAEPEVP QVETKEVQLAEAEVLLAKVTDSSLSKANATETLAGLRNLLTLQIMDNNSIMAEAKL LALLKGSNPFSSVSSEKIN"</div>															
ORIGIN																
Query Match	100.0%; Score 2389; DB 1; Length 2541;															
Best Local Similarity	100.0%; Pred. No. 0;															
Matches 2389;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;															
QY	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA	60													
Db	150	TTCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTA	209													
QY	61	TATAGATGGAATAACAGCGACGCAAAAAACGGAGAAATTTCGCTCTGATGAGGTTAGCAA	120													
Db	210	TATAGATGGAATAACAGCGACGCAAAAAACGGAGAAATTTCGCTCTGATGAGGTTAGCAA	269													
QY	121	CGGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC	180													
Db	270	CGGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC	329													
QY	181	TTCCATGCGCGACCACTATCATTTATCAATGGTAAGGTTCCCTTATGAGCTATCATCAG	240													
Db	330	TTCCATGCGCGACCACTATCATTTATCAATGGTAAGGTTCCCTTATGAGCTATCATCAG	389													
QY	241	TGAAGAAATTACTCATGAAGATCCAAACTAATAGCTAAAGATGAGGATATTGTTAATGA	300													
Db	390	TGAAGAAATTACTCATGAAGATCCAAACTAATAGCTAAAGATGAGGATATTGTTAATGA	449													
QY	301	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGGATGC	360													
Db	450	GGTCAAGGGTGGATATGTTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGGATGC	509													
QY	361	TGCCACCGCGATAACGTCGCTACAAAGAGGAAATCAATCGCAAAAAACAAGAGCATAG	420													
Db	510	TGCCACCGCGATAACGTCGCTACAAAGAGGAAATCAATCGCAAAAAACAAGAGCATAG	569													
QY	421	TCAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGCTTGGCTTGGCAGCTTCGCA	480													
Db	570	TCAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGCTTGGCTTGGCAGCTTCGCA	629													
QY	481	AGGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC	540													
Db	630	AGGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC	689													
QY	541	TGGTGTATGCTTTATATCGTTCTCATGAGATCATTAACATTACATTCCTTAAGAAATGAGTT	600													
Db	690	TGGTGTATGCTTTATATCGTTCTCATGAGATCATTAACATTACATTCCTTAAGAAATGAGTT	749													
QY	601	ATCAGCTAGCGAGTTGGTGTGCGAAGCCTTCTTATCTGGTTCGAGGAAATCTGTCAA	660													
Db	750	ATCAGCTAGCGAGTTGGTGTGCGAAGCCTTCTTATCTGGTTCGAGGAAATCTGTCAA	809													

QY	661	TTCAAGAACTATCGCGACAAATAAGCGATAACACTTCAAGAACAAACTGGGTACCTTC	720
DB	810	TTCAAGAACTATCGCGACAAATAAGCGATAACACTTCAAGAACAAACTGGGTACCTTC	869
QY	721	TGTAAGCAATCCAGGAACTACAAATACTAACCAAGCAACCAAGCAAACTAACAGTCA	780
DB	870	TGTAAGCAATCCAGGAACTACAAATACTAACCAAGCAACCAAGCAAACTAACAGTCA	929
QY	781	AGCAAGTCAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAACTGCTTTGAG	840
DB	930	AGCAAGTCAAAGTAATGACATTTGATAGTCTCTTTGAAACAGCTCTACAACTGCTTTGAG	989
QY	841	TCAACACACATGTAGATCTGATGGCCCTCTCTTTGATCCAGCACAATCACAGTCAAC	900
DB	990	TCAACACACATGTAGATCTGATGGCCCTCTCTTTGATCCAGCACAATCACAGTCAAC	1049
QY	901	AGCTAGAGTGTTCAGTGCACACAGAGATCATTAACCACTTCATCCCTTACTCTCAAAAT	960
DB	1050	AGCTAGAGTGTTCAGTGCACACAGAGATCATTAACCACTTCATCCCTTACTCTCAAAAT	1109
QY	961	GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATTCCTCTTTCGTTTCGTTCAAACTTG	1020
DB	1110	GTCTGAATTTGGAAGAACGAATCGCTCGTATTTATTCCTCTTTCGTTTCGTTCAAACTTG	1169
QY	1021	GGTACCAGATTCAAGGCCAGAAACCAAGTCCACAACCGACTCCCGGAACTAGTCCAGG	1080
DB	1170	GGTACCAGATTCAAGGCCAGAAACCAAGTCCACAACCGACTCCCGGAACTAGTCCAGG	1229
QY	1081	CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTCGTTTCGTTTCGTT	1140
DB	1230	CCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTCGTTTCGTTTCGTT	1289
QY	1141	AGCAAAAGTTGGGAAAGGATATGATTCGAAGAAAGGGAATCTCTCGTTATGCTTTGTC	1200
DB	1290	AGCAAAAGTTGGGAAAGGATATGATTCGAAGAAAGGGAATCTCTCGTTATGCTTTGTC	1349
QY	1201	GAAGATTTTACCCTCTGAACTCTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG	1260
DB	1350	GAAGATTTTACCCTCTGAACTCTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG	1409
QY	1261	TGTTTTCACACACTTTAACTGTCTTAAAGAAATGTTGCTCTCTGAGCAAGAAATTTTA	1320
DB	1410	TGTTTTCACACACTTTAACTGTCTTAAAGAAATGTTGCTCTCTGAGCAAGAAATTTTA	1469
QY	1321	TGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGGCTTGTGTTGNAATAAAGGCTG	1380
DB	1470	TGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGGCTTGTGTTGNAATAAAGGCTG	1529
QY	1381	TAATTTCTGATTTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA	1440
DB	1530	TAATTTCTGATTTTCAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA	1589
QY	1441	TAAAGAAAATTTGGTAGATGATTTTATTTGGCAATTCCTAGCACCAATTTACCCATCCAGAGCG	1500
DB	1590	TAAAGAAAATTTGGTAGATGATTTTATTTGGCAATTCCTAGCACCAATTTACCCATCCAGAGCG	1649
QY	1501	ACTTGGCAAAACCAATTTCTCAAAATTTGATGATGATGATGATGATGATGATGATGATGAT	1560
DB	1650	ACTTGGCAAAACCAATTTCTCAAAATTTGATGATGATGATGATGATGATGATGATGATGAT	1709
QY	1561	AGCTGATAGTATACAAAGTGGTGTGATGATTTTATTTGATGAACATGATATATCATGCTCA	1620
DB	1710	AGCTGATAGTATACAAAGTGGTGTGATGATTTTATTTGATGAACATGATATATCATGCTCA	1769
QY	1621	TGAAGGAGATGATATGTAACCCCTCATATGGCCCATAGTCTACTGATTTGGAAGAGATAG	1680
DB	1770	TGAAGGAGATGATATGTAACCCCTCATATGGCCCATAGTCTACTGATTTGGAAGAGATAG	1829
QY	1681	CTTTCTCATAGGAAAGTGGAGCTCAAGCCTTATCTAAGAAAGAGGATGATCTCTACC	1740
DB	1830	CTTTCTCATAGGAAAGTGGAGCTCAAGCCTTATCTAAGAAAGAGGATGATCTCTACC	1889

Db 780 TGTAAAGCAATCCAGGAACCTACAAATATCTAAACAGCAACCAACAGCACTAACAGTCA 839  
Qy 781 AGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAAAGTGCCTTTGAG 840  
Db 840 AGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAAAGTGCCTTTGAG 899  
Qy 841 TCAACGACATGTAGAACTGATGCTGCTTTGATCCAGCAAAATCAAAATCAAGTCAAC 900  
Db 900 TCAACGACATGTAGAACTGATGCTGCTTTGATCCAGCAAAATCAAAATCAAGTCAAC 959  
Qy 901 AGCTAGAGGTCTGAGTGCACACGAGATCAATACCACTTTCACTCTCTCAAT 960  
Db 960 AGCTAGAGGTCTGAGTGCACACGAGATCAATACCACTTTCACTCTCTCAAT 1019  
Qy 961 GTCTGAATTTGAAGAAAGCAATGCTGCTGATTTATCCCTCTGTTATCGTTCAAAACCAATG 1020  
Db 1020 GTCTGAATTTGAAGAAAGCAATGCTGCTGATTTATCCCTCTGTTATCGTTCAAAACCAATG 1079  
Qy 1021 GGTACAGATTCAGGCGAGCAACCAAGTCCACACCGACTCCGAAACCTAGTCCAGG 1080  
Db 1080 GGTACAGATTCAGGCGAGCAACCAAGTCCACACCGACTCCGAAACCTAGTCCAGG 1139  
Qy 1081 CCCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCACTGGT 1140  
Db 1140 CCCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCACTGGT 1199  
Qy 1141 ACGAAAGTTGGGGAAGATATGATTCGAAGAAAGGCAATCTCTGTTATGCTTTG 1200  
Db 1200 ACGAAAGTTGGGGAAGATATGATTCGAAGAAAGGCAATCTCTGTTATGCTTTG 1259  
Qy 1201 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 1260  
Db 1260 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAG 1319  
Qy 1261 TGTTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320  
Db 1320 TGTTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1379  
Qy 1321 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTGTTGTTGTTGTTGTTG 1380  
Db 1380 TGATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCCTGTTGTTGTTGTTGTTGTTG 1439  
Qy 1381 TAATCTGATTTCCAGCTTAGCAATATTAAGACGCTTGAATGATGATGATGATGATGATGAT 1440  
Db 1440 TAATCTGATTTCCAGCTTAGCAATATTAAGACGCTTGAATGATGATGATGATGATGATGAT 1499  
Qy 1441 TAAAGAAAAATTTGTTAGATGATTTATTTGGCAATCTCTAGCAACCAATTTACCCATCCAGAGCG 1500  
Db 1500 TAAAGAAAAATTTGTTAGATGATTTATTTGGCAATCTCTAGCAACCAATTTACCCATCCAGAGCG 1559  
Qy 1501 ACTTGGCAACCAATTTCTAAATGATGATTAAGAGCAAGTTCGTTATGCTCAAT 1560  
Db 1560 ACTTGGCAACCAATTTCTAAATGATGATTAAGAGCAAGTTCGTTATGCTCAAT 1619  
Qy 1561 AGCTGATAAGTATCAAGCTCAGATGTTACATTTTGGTGAACATGATATATATCAAGTGA 1620  
Db 1620 AGCTGATAAGTATCAAGCTCAGATGTTACATTTTGGTGAACATGATATATATCAAGTGA 1679  
Qy 1621 TGAAGGAGATCATATGTAAAGCTCATATGAGGCAATGATGATGATGATGATGATGATGATGAT 1680  
Db 1680 TGAAGGAGATCATATGTAAAGCTCATATGAGGCAATGATGATGATGATGATGATGATGATGAT 1739  
Qy 1681 CCTTCTGATAGGAAAGTTGAGCTCAAGCTTACTAAAGAAAGTATCTTACC 1740  
Db 1740 CCTTCTGATAGGAAAGTTGAGCTCAAGCTTACTAAAGAAAGTATCTTACC 1799  
Qy 1741 TCCATCTCCAGCAGATGTTAAAGCAATCCCACTGGATAGTATGATGATGATGATGATGATGAT 1800  
Db 1800 TCCATCTCCAGCAGATGTTAAAGCAATCCCACTGGATAGTATGATGATGATGATGATGATGAT 1859  
Qy 1801 CAATCGTGAAGGGAAGAAACGAATTCCTGCTTCCATATATGTTGAGCA 1860

Db 1860 CAATCGTGTGAAGGGGAAAAACGAATTCACATCGTTCCGACTCCATATATGTTGAGCA 1919  
Qy 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATATTTCTCTATAAGATCAATTTACCATATAT 1920  
Db 1920 TACAGTTGAGGTTAAAAACGGTAATTTGATATTTCTCTATAAGATCAATTTACCATATAT 1979  
Qy 1921 TAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 1980  
Db 1980 TAAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGA 2039  
Qy 1981 TTTGTTTGCACGATTAAGTACTACGTAGAACACCTCTGAGGAAGCTCCACATTTCAATGA 2040  
Db 2040 TTTGTTTGCACGATTAAGTACTACGTAGAACACCTCTGAGGAAGCTCCACATTTCAATGA 2099  
Qy 2041 TGGATGGGCAATGCTCAGTGAAGCATGTTGTTAGCAAGAAAGACACAGTGAAGATCCAAA 2100  
Db 2100 TGGATGGGCAATGCTCAGTGAAGCATGTTGTTAGCAAGAAAGACACAGTGAAGATCCAAA 2159  
Qy 2101 TAAAGACTTCAAGAGCGATGAAGAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCC 2160  
Db 2160 TAAAGACTTCAAGAGCGATGAAGAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCC 2219  
Qy 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAAGTCTTGTTCGGAA 2220  
Db 2220 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAAGTCTTGTTCGGAA 2279  
Qy 2221 AGTAACGGATTTAGTCTGAAAGCCATGCAACAGAAACTCTAGCTGTTTACGAATAA 2280  
Db 2280 AGTAACGGATTTAGTCTGAAAGCCATGCAACAGAAACTCTAGCTGTTTACGAATAA 2339  
Qy 2281 TTTGACTCTTCAATATGATTAACATAGTATCATGCGAAGACAGAAAAATTTACTTGC 2340  
Db 2340 TTTGACTCTTCAATATGATTAACATAGTATCATGCGAAGACAGAAAAATTTACTTGC 2399  
Qy 2341 GTTGTAAAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2389  
Db 2400 GTTGTAAAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 6  
AF291695 2541 bp DNA linear BCT 15-MAR-2001  
LOCUS  
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad A protein  
(phtA) gene, complete cds.  
ACCESSION AF291695  
VERSION AF291695.1 GI:13345012  
KEYWORDS  
SOURCE Streptococcus pneumoniae  
ORGANISM Streptococcus pneumoniae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 2541)  
AUTHORS Wizenann,T.M., Heinrichs,J.H., Adamou,J.E., Erwin,A.L., Kunsch,C.,  
Choi,G.H., Barash,S.C., Rosen,C.A., Masure,H.R., Tuomanen,E.,  
Gayle,A., Brewah,Y.A., Walsh,W., Barren,P., Lathigra,R., Hanson,M.,  
Langermann,S., Johnson,S. and Koenig,S.  
TITLE Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection  
JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)  
MEDLINE 21116976  
PUBMED 11179332  
REFERENCE 2 (bases 1 to 2541)  
AUTHORS Choi,G.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,  
Inc., 9410 Key West Ave., Rockville, MD 20850, USA  
FEATURES  
source Location/Qualifiers  
1..2541  
/organism="Streptococcus pneumoniae"  
/mol\_type="genomic DNA"  
/strain="N4"  
/db\_xref="taxon:1313"  
91..2541  
gene







QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAGCGTTAAGGAAATAATCGTGTTCCTTA 60  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAGCGTTAAGGAAATAATCGTGTTCCTTA 60  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 61 TATAGATGGAAGAAACAGCGACGCGAAGAAACCGAGAAATTTGACTCTGATGAGGTTAGCAA 120  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 61 TATAGATGGAAGAAACAGCGACGCGAAGAAACCGAGAAATTTGACTCTGATGAGGTTAGCAA 120  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 121 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGCGCTATGTCAC 180  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 121 GCGTGAAGGAATCAATGCTCAGCAAAATCGTCATCAAGATAACAGCAAGCGCTATGTCAC 180  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 181 TTCACATGCGGACCACTATCATTAATTAACAATGGTAAGGTTCCCTATGACGCTATCATCAG 240  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 241 TGAAGAAATTAATCAATGAAAGATCCAACTAATAAGCTAAAGATGAGGATATGTTAATGA 300  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 241 TGAAGAAATTAATCAATGAAAGATCCAACTAATAAGCTAAAGATGAGGATATGTTAATGA 300  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 301 GGTCAAGGTTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGC 360  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 361 TGCCCAACGCGGATAACGTCGGTACAAAAGAGGAAATCAATCGACAAAACAAAGAGCATAG 420  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGGCTTGGCAGCTTCGCA 480  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 421 TCAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGTTGGCTTGGCAGCTTCGCA 480  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 481 AGGACCTATACATACAGATGATGTTATATCTTTAATGTTCTGATATCATAGAGATAC 540  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 481 AGGACCTATACATACAGATGATGTTATATCTTTAATGTTCTGATATCATAGAGATAC 540  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 541 TGTGTATGCTTATATCGTTTCTCATGAGATCATTAACCATTAATTCCTAAGAAATGAGTT 600  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 541 TGTGTATGCTTATATCGTTTCTCATGAGATCATTAACCATTAATTCCTAAGAAATGAGTT 600  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 601 ATCAGTACGAGTTGGCTGCTGCAGAACGCTTCCTATCTGCTGAGGAAATCTGTCAAA 660  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 601 ATCAGTACGAGTTGGCTGCTGCAGAACGCTTCCTATCTGCTGAGGAAATCTGTCAAA 660  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 661 TTCAAGAACTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 661 TTCAAGAACTATCGCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTC 720  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 721 TGTAAAGCAATCCAGGAATACAAAATACTAAACAAGCAACAAAGCAACAACTAACAGTCA 780  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 721 TGTAAAGCAATCCAGGAATACAAAATACTAAACAAGCAACAAAGCAACAACTAACAGTCA 780  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 781 AGCAAGTCAAAGTAATGATGATGCTCTTGAACACAGCTCTACAAAATCGCTTTGAG 840  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 781 AGCAAGTCAAAGTAATGATGATGCTCTTGAACACAGCTCTACAAAATCGCTTTGAG 840  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 841 TCAACGACATGATGATGCTCTTGAACACAGCTCTACAAAATCGCTTTGAG 900  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 841 TCAACGACATGATGATGCTCTTGAACACAGCTCTACAAAATCGCTTTGAG 900  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 901 AGCTAGAGTGTGAGTGCACAGGATCATTAACCACTTATCCCTTACTCTCAAT 960  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 901 AGCTAGAGTGTGAGTGCACAGGATCATTAACCACTTATCCCTTACTCTCAAT 960  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 961 GTCTCAATTTGAAGAACGAACTCGCTGATTAATTCCTTCTGTTATCGTTCAACCATTTG 1020  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 961 GTCTCAATTTGAAGAACGAACTCGCTGATTAATTCCTTCTGTTATCGTTCAACCATTTG 1020  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1021 GGTACCAATTTCAAGCCAGAACAAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG 1080  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1021 GGTACCAATTTCAAGCCAGAACAAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGG 1080  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1081 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGGT 1140  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1081 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGGT 1140  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1141 ACAGAAAGTTGGGAAGGATATGTTATTCGAAGAAAGGGCATCTCTGTTATGCTTTTGC 1200  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1141 ACAGAAAGTTGGGAAGGATATGTTATTCGAAGAAAGGGCATCTCTGTTATGCTTTTGC 1200  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1201 GAAGAAATTTACCATCTGAAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1201 GAAGAAATTTACCATCTGAAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1261 TGTTCACACATCTTTAACTGCTTAAAGAAAGAAATGTTGCTCCTGTCGACCAAGAAATTTA 1320  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1261 TGTTCACACATCTTTAACTGCTTAAAGAAAGAAATGTTGCTCCTGTCGACCAAGAAATTTA 1320  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGNAATTAAGGCTG 1380  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1321 TGATAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTTGTGTTGNAATTAAGGCTG 1380  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1381 TAATCTGATTTCCAGCCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1381 TAATCTGATTTCCAGCCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAA 1440  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1441 TAAAGAAATTTGCTAGATGATTTATTTGGCATTTCTAGACCAATTAACCATCCAGAGCG 1500  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1441 TAAAGAAATTTGCTAGATGATTTATTTGGCATTTCTAGACCAATTAACCATCCAGAGCG 1500  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1501 ACTTGGCAAAACCAATTTCTCAATGAGTATATCTGAAGACGAAAGTTCGTAATTCCTCAAT 1560  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1501 ACTTGGCAAAACCAATTTCTCAATGAGTATATCTGAAGACGAAAGTTCGTAATTCCTCAAT 1560  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1561 AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGTAACATGATATATCATGTA 1620  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1561 AGCTGATAAGTATACAAACGTCAGATGTTTACATTTTGTGTAACATGATATATCATGTA 1620  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1621 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGATTTGGAAGAAAGTAG 1680  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1621 TGAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGATTTGGAAGAAAGTAG 1680  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1681 CTTTCTGATAGGAAAGTGTGAGCTCAAGCTATATACTAAAGAAAGAAAGTATCTTACC 1740  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1681 CTTTCTGATAGGAAAGTGTGAGCTCAAGCTATATACTAAAGAAAGAAAGTATCTTACC 1740  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1741 TCCATCTCCAGACGCGATGTTAAAGCAATCCACTGAGGATAGTGCAGCAGCTATTTA 1800  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1741 TCCATCTCCAGACGCGATGTTAAAGCAATCCACTGAGGATAGTGCAGCAGCTATTTA 1800  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTGCTTCCACTTCCATATATGTTGAGCA 1860  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1801 CAATCGTGTGAAAGGGGAAAAACGAATTCACCTGCTTCCACTTCCATATATGTTGAGCA 1860  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCTAAGGATCATTAACATATAT 1920  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCTAAGGATCATTAACATATAT 1920  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1921 TAAATTTGCTGTTTGTGATGATCACAATCAAGCTCCAAATGGCTATACCTTGAAGA 1980  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1921 TAAATTTGCTGTTTGTGATGATCACAATCAAGCTCCAAATGGCTATACCTTGAAGA 1980  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1981 TTTGTTGCGACGATTAAGTACTAGTAGAACCCCTGACGAACCTCCACATTTCTAATGA 2040  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 1981 TTTGTTGCGACGATTAAGTACTAGTAGAACCCCTGACGAACCTCCACATTTCTAATGA 2040  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 2041 TGGATGGGCAATGTCAGTGCATGTTAGGCAAGAAACCAAGCAAGTGAAGATCCAAA 2100  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 2041 TGGATGGGCAATGTCAGTGCATGTTAGGCAAGAAACCAAGCAAGTGAAGATCCAAA 2100  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 2101 TAAAGAACTTCAAAGCGGATGAAGCCAGTACAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 2101 TAAAGAACTTCAAAGCGGATGAAGCCAGTACAGGAAACACCTGCTGAGCCAGAGTCCC 2160  
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 QY 2161 TCAAGTAGAGCTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTTCGTTGCGAA 2220



```
QY 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAACTCGCTTTGAG 840
Db 781 AGCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAACTCGCTTTGAG 840
QY 841 TCAACGACATGTAGAACTCTGATGGCTTCTCTTTGATCCAGACAAATCACAAAGTCGAAC 900
Db 841 TCAACGACATGTAGAACTCTGATGGCTTCTCTTTGATCCAGACAAATCACAAAGTCGAAC 900
QY 901 AGCTAGAGTGTTCAGTCCACACGAGATCATTTACCACTTCATCCCTTATCTCAAAAT 960
Db 901 AGCTAGAGTGTTCAGTCCACACGAGATCATTTACCACTTCATCCCTTATCTCAAAAT 960
QY 961 GTCGTAATTTGGAAGACGAATCGCTCGTATTATTCCTTTCGTTATCGTTTCAAAACATTG 1020
Db 961 GTCGTAATTTGGAAGACGAATCGCTCGTATTATTCCTTTCGTTATCGTTTCAAAACATTG 1020
QY 1021 GGTACCAAGTTCAAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACTTAGTCCAGG 1080
Db 1021 GGTACCAAGTTCAAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACTTAGTCCAGG 1080
QY 1081 CCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTCTTCTTGGTTAGTCAGTGGT 1140
Db 1081 CCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTCTTCTTGGTTAGTCAGTGGT 1140
QY 1141 ACGAAAAGTTGGGAAGGATATGTATTGCAAGAAAGGGCATCTCTCGTTATGTCTTTGC 1200
Db 1141 ACGAAAAGTTGGGAAGGATATGTATTGCAAGAAAGGGCATCTCTCGTTATGTCTTTGC 1200
QY 1201 GAAAGATTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
Db 1201 GAAAGATTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
QY 1261 TGTTTCACACACTTAACTGCTTAAAAAGAAAATGTTGCTCTCTCGTGACCAAGAAATTTTA 1320
Db 1261 TGTTTCACACACTTAACTGCTTAAAAAGAAAATGTTGCTCTCTCGTGACCAAGAAATTTTA 1320
QY 1321 TGATAAGCATATACTCTGTAACAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCG 1380
Db 1321 TGATAAGCATATACTCTGTAACAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCG 1380
QY 1381 TAATTCGTGATTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440
Db 1381 TAATTCGTGATTTCCAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAA 1440
QY 1441 TAAAGAAAATTTGGTAGATGATTTATGSCATTCCTAGCACCAATTTACCATCCAGAGCG 1500
Db 1441 TAAAGAAAATTTGGTAGATGATTTATGSCATTCCTAGCACCAATTTACCATCCAGAGCG 1500
QY 1501 ACTTGGCAACCAAAATTTCTCAAAATTCAGTATACTGAAGACGAAGTTGCTATTGCTCAATT 1560
Db 1501 ACTTGGCAACCAAAATTTCTCAAAATTCAGTATACTGAAGACGAAGTTGCTATTGCTCAATT 1560
QY 1561 AGCTGATAAGTATACAACTCAGATGTTTACATTTTGTGTAACATGATATATATCAAGTGA 1620
Db 1561 AGCTGATAAGTATACAACTCAGATGTTTACATTTTGTGTAACATGATATATATCAAGTGA 1620
QY 1621 TGAAGGAGTGCATATGTAACGCTCATATGGCCATAGTCACTGATGTTGGAAGAGATAG 1680
Db 1621 TGAAGGAGTGCATATGTAACGCTCATATGGCCATAGTCACTGATGTTGGAAGAGATAG 1680
QY 1681 CTTTCTGATAAGGAAAAGTTGACGCTCAAGCCTATCTAAAGAAAAAGGTATCTTACC 1740
Db 1681 CTTTCTGATAAGGAAAAGTTGACGCTCAAGCCTATCTAAAGAAAAAGGTATCTTACC 1740
QY 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTTA 1800
Db 1741 TCCATCTCCAGACGAGATGTTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTTA 1800
QY 1801 CAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860
Db 1801 CAATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTTCGACTTCCATATATGTTGAGCA 1860
```

```
QY 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCATTAACCAATAT 1920
Db 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAGGATCATTAACCAATAT 1920
QY 1921 TAAATTTCTCTTGGTTTGTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 1921 TAAATTTCTCTTGGTTTGTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
QY 1981 TTTGTTTCGGACGATTAAGTACTAGTAGAACACACCTGACGAACGCTCCACATTTCTAATGA 2040
Db 1981 TTTGTTTCGGACGATTAAGTACTAGTAGAACACACCTGACGAACGCTCCACATTTCTAATGA 2040
QY 2041 TCGATGGGGCAATGCCAGTGAAGCATGTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
Db 2041 TCGATGGGGCAATGCCAGTGAAGCATGTGTAGGCAAGAAAGACCAAGTGAAGATCCAAA 2100
QY 2101 TAAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAAACACCTGCTGAGCCAGAAAGTCCC 2160
Db 2101 TAAAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAAACACCTGCTGAGCCAGAAAGTCCC 2160
QY 2161 TCAAGTAGAGACTGAAAAGTAGAGCCCAACTCAAGNAGCAGAGTTTTCCTTGGCGAA 2220
Db 2161 TCAAGTAGAGACTGAAAAGTAGAGCCCAACTCAAGNAGCAGAGTTTTCCTTGGCGAA 2220
QY 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
Db 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 2280
QY 2281 TTTGACTCTTCAAAATTTAGGATAACAATAGTATCATGTCAGAGAACGAGAAAATTTACTTGC 2340
Db 2281 TTTGACTCTTCAAAATTTAGGATAACAATAGTATCATGTCAGAGAACGAGAAAATTTACTTGC 2340
QY 2341 GTTGTAAAGGAAGTAATCTCTCATCTGTAGTAAAGGAAAAAATAAAC 2389
Db 2341 GTTGTAAAGGAAGTAATCTCTCATCTGTAGTAAAGGAAAAAATAAAC 2389
```

## RESULT 3

BD063274 2389 bp DNA linear PAT 27-AUG-2002  
Streptococcus pneumoniae antigens and vaccines.

BD063274  
BD063274.1 GI:22608877  
JP 2001505415-A/28.

KEYWORDS  
JP 2001505415-A/28.  
SOURCE  
unidentified  
ORGANISM  
unclassified.

REFERENCE  
1 (bases 1 to 2389)

AUTHORS  
Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromockyj,A.  
TITLE  
Streptococcus pneumoniae antigens and vaccines  
JOURNAL  
Patent: JP 2001505415-A 28 24-APR-2001;  
HUMAN GENOME SCIENCES INC

COMMENT  
FN JP 2001505415-A/28  
PD 24-APR-2001  
PF 30-OCT-1997 JP 1998520667  
PR 31-OCT-1996 US 60/029960

PI CHARLES A KUNSCH,GIL H CHOI,SYDNOR L JOHNSON,ALEX HROMOCKYJ PC  
C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC  
G01N33/569,  
PC G01N33/68  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers.

FEATURES  
source

1..2389  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 100.0%; Score 2389; DB 6; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



66	24	1.0	1342	6	AR341019	Sequence
67	24	1.0	1342	6	BD063337	Streptoco
68	24	1.0	1446	6	AX608407	Sequence
69	24	1.0	1455	6	BD263587	Novel str
70	24	1.0	1455	6	BD264131	Streptoco
71	24	1.0	1455	6	BD268047	Streptoco
72	24	1.0	1455	6	AR344446	Sequence
73	24	1.0	1455	6	BD223895	NuclLeic a
74	24	1.0	2528	6	BD263592	Novel str
75	24	1.0	2528	6	AX343078	Sequence
76	24	1.0	3117	6	AX568780	Sequence
77	24	1.0	3120	1	AF318956	Streptoco
78	24	1.0	3120	6	BD263584	Novel str
79	24	1.0	3120	6	AX343070	Sequence
80	24	1.0	5048	6	BD263589	Novel str
81	24	1.0	5048	6	AX343071	Sequence
82	24	1.0	6867	6	AR218960	Sequence
83	24	1.0	6867	6	BD003872	Polynucle
84	24	1.0	21397	1	AE014279	Streptoco
85	24	1.0	75248	6	AX602204	Sequence
86	24	1.0	98050	1	SAG766854	Streptoco
87	23	1.0	154101	2	AC136885	Sus scrof
88	23	1.0	194474	2	AC137536	Sus scrof
89	23	1.0	203434	10	AL732478	Mouse DNA
90	23	1.0	214633	2	AC115959	Mus muscu
ALIGNMENTS						
RESULT 1	AR120265	2389 bp	DNA	linear	PAT 16-MAY-2001	
LOCUS	Sequence 55 from patent US 6159469.					
DEFINITION	AR120265					
ACCESSION	AR120265.1	GI:14103841				
VERSION						
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 2389)					
AUTHORS	Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A.					
TITLE	Streptococcus pneumoniae antigens and vaccines					
JOURNAL	Patent: US 6159469-A 55 12-DEC-2000;					
FEATURES	Location/Qualifiers					
source	1..2389					
	/organism="unknown"					
	/mol_type="unassigned DNA"					
ORIGIN						
Query Match	100.0%; Score 2389; DB 6; Length 2389;					
Best Local Similarity	100.0%; Pred. No. 0;					
Matches 2389; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA	60			
Db	1	TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTA	60			
QY	61	TATAGATGAAACACAGCCGCAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA	120			
Db	61	TATAGATGAAACACAGCCGCAAAACCGAGAAATTTGACTCCTGATGAGTTAGCAA	120			
QY	121	GCGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGGCTATGTAC	180			
Db	121	GCGTGAAGGANCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGGCTATGTAC	180			
QY	181	TTCAATGGCGACCATATCAATTTAATGGTAAGTTTCCTTATGACGCTATCAATCAG	240			
Db	181	TTCAATGGCGACCATATCAATTTAATGGTAAGTTTCCTTATGACGCTATCAATCAG	240			
QY	241	TGAGATTTACTCATGAAAGTCCAACTATAGCTTAAAGATGAGGATATTGTTAATGA	300			

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2004, 22:17:38 ; Search time 9275 Seconds  
(without alignments)  
11164.044 Million cell updates/sec

Title: US-09-765-271-55  
Perfect score: 2389  
Sequence: 1 TTCTACAGTTGGACTGT.....TAAGTAAGGAAAAATAAAC 2389

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rtd.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2389	100.0	2389	6	AR120265	AR120265 Sequence
2	2389	100.0	2389	6	AR340956	AR340956 Sequence
3	2389	100.0	2389	6	BD063274	BD063274 Streptoco
4	2389	100.0	2451	6	BD268048	BD268048 Streptoco
5	2389	100.0	2451	6	AR344447	AR344447 Sequence
6	2389	100.0	2541	1	AF291695	AF291695 Streptoco
7	2338	97.9	2406	6	AX569139	AX569139 Sequence
8	2338	97.9	8195	6	AR218862	AR218862 Sequence
9	2338	97.9	8195	6	BD003774	BD003774 Polynucle
c 10	2338	97.9	10256	1	AE007418	AE007418 Streptoco
c 11	2338	97.9	349980	6	AX571763	AX571763 Sequence
c 12	1624	68.0	10320	1	AE008479	AE008479 Streptoco
13	927	38.8	20035	2	SPNEU1915	AL449937 Streptoco
14	326	13.6	702	2	SPNEU1929	AL449951 Streptoco
15	200	8.4	2535	1	AF340221	AF340221 Streptoco
16	148	6.2	2523	6	BD263585	BD263585 Novel str
17	148	6.2	2523	6	AX343072	AX343072 Sequence
18	148	6.2	2647	6	BD263590	BD263590 Novel str
19	148	6.2	2647	6	AX343073	AX343073 Sequence
20	53	2.2	504	6	BD229970	BD229970 Human com
c 21	53	2.2	2166	12	AF340222	AF340222 Synthetic
22	53	2.2	2290	6	AR120270	AR120270 Sequence
23	53	2.2	2290	6	AR340961	AR340961 Sequence
24	53	2.2	2290	6	BD063279	BD063279 Streptoco
25	53	2.2	2359	6	AR219011	AR219011 Sequence
26	53	2.2	2359	6	BD003923	BD003923 Polynucle
27	53	2.2	2457	1	AF318954	AF318954 Streptoco
28	53	2.2	2457	6	AX569137	AX569137 Sequence
29	53	2.2	2478	6	BD229972	BD229972 Human com
30	53	2.2	2481	6	BD264091	BD264091 Streptoco
31	53	2.2	2517	1	AF318955	AF318955 Streptoco
32	53	2.2	2517	6	AX568778	AX568778 Sequence
33	53	2.2	2531	6	BD268046	BD268046 Streptoco
34	53	2.2	2531	6	BD268049	BD268049 Streptoco
35	53	2.2	2531	6	AR344445	AR344445 Sequence
36	53	2.2	2531	6	AR344448	AR344448 Sequence
37	53	2.2	2639	6	BD263591	BD263591 Novel str
38	53	2.2	2639	6	AX343074	AX343074 Sequence
39	53	2.2	11931	1	AE007403	AE007403 Streptoco
40	53	2.2	12372	1	AE008464	AE008464 Streptoco
41	53	2.2	12372	2	SPNEU1907	AL449929 Streptoco
42	53	2.2	75874	2	SPNEU1901	AL449923 Streptoco
43	53	2.2	232807	6	AX571762	AX571762 Sequence
44	53	2.2	349980	6	AX571762	AF340223 Synthetic
45	50	2.1	492	12	AF340223	AR219123 Sequence
46	44	1.8	973	6	AR219123	BD004035 Polynucle
47	44	1.8	973	6	BD004035	BD263631 Novel str
48	38	1.6	3171	6	BD263631	AR120406 Sequence
49	30	1.3	40	6	AR120406	BD268045 Streptoco
c 50	30	1.3	40	6	BD268045	AR341097 Sequence
c 51	30	1.3	40	6	AR344444	AR344444 Sequence
c 52	30	1.3	40	6	AR344444	BD063415 Streptoco
c 53	30	1.3	36	6	BD268043	BD268043 Streptoco
54	27	1.1	36	6	AR344442	AR344442 Sequence
55	27	1.1	36	6	AR120405	AR120405 Sequence
56	27	1.1	37	6	AR341096	BD063414 Streptoco
57	27	1.1	37	6	BD063414	BD268044 Streptoco
58	27	1.1	35	6	AR344443	AR344443 Sequence
59	26	1.1	35	6	BD263613	BD263613 Novel str
60	26	1.1	33	6	AX343114	AX343114 Sequence
61	25	1.0	33	6	BD263615	BD263615 Novel str
62	25	1.0	34	6	AX343116	AX343116 Sequence
63	25	1.0	34	6	AR120328	AR120328 Sequence
64	25	1.0	1342	6	AR120328	

Qy	562	SDKEKVAQAQYTK	EGKGLPPSPDADVKANPTGDSAAAIYNRVKGEKRI	ELVRLPYMVEHT	621
Db	302	SEAEARAQAQAKE	GLTPTPDHDSQNGTEAKGAEAIYNRVKA	AKKVFDRMPYNLQYT	361
Qy	622	VEVKNGLII	PHKDYHNIIKFAWFDHTVKA	NGYTLDFATIIKYVVEHPDERPHSNDG	681
Db	362	VEVKNGLII	PHYDRYHNIIKFEWDEGLYEAPKGYTL	DLATVKYVVEHPDERPHSNDG	421
Qy	682	WGNASEHVGLK	KHSDENPNKFKAD	-----EPEVE	711
Db	422	FGNASDHVQRN	KNGQADNTQEKPSSEKPTQEKPREETPREEK	POSEKPESEKPTPEEPE	481
Qy	712	ETPAEPEVPOVET	KEVQAQAEVLLAKVTDSSIIKANATETLAGIRN	NLTQIOMDNNSI	771
Db	482	ESPRESPEVET	KEVBEKREAEOLLGKIQPIIKSNAKETLTGL	KNLLFGTQDNNTI	541
Qy	772	MAAEKULALIKGS	-----NFSVS	791	
Db	542	MAAEKLIALIK	ESKGNPMOP	SOLS	566

## RESULT 25

```

US-10-324-143-26
; Sequence 26, Application US/10324143
; Publication NO. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHABLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 555
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-26

```

[illegible]



```
QY 774 EAEKLLALLKGS 785
Db 601 EAEKLLALLKGS 612

RESULT 21
US-09-884-465A-383
; Sequence 383, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 383
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (557)-(557)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (558)-(558)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-383

Query Match 45.6%; Score 1899; DB 10; Length 1126;
Best Local Similarity 64.9%; Pred. No. 8.5e-128;
Matches 366; Conservative 63; Mismatches 101; Indels 34; Gaps 3;

QY 266 DIDLKQLYKLPISQRHVESDGLVDFPAQITTSRTARGVAVPHGHHYHFIYEQMSELEE 325
Db 2 DIDLKQLYKLPISQRHVESDGLVDFPAQITTSRTARGVAVPHGHHYHFIYEQMSELEE 61

QY 326 RIARIIPLYRSNHWVPSRPEOPSPOPTPEPSGPQAPNLK-IDSN--SSLVSQLVR 381
Db 62 RIARIIPLYRSNHWVPSRPEOPSPOPTPEPSGPQAPNPQAPNPIDKLVKEAVR 121

QY 382 KVGEGYVFEERKIGSYVFAKDLPSFTVKNLESKLSKQESVSHLTLTAKKENVAPRDQEFYD 441
Db 122 KVGEGYVFEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDRFYN 181

QY 442 KAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESNKEKLVDDLLAFAPITHPERL 501
Db 182 KAYDILLARIHQDLLDNKGRQVDFEALDNLLERLKDVSDDKVLVDDIILAFAPIRHPERL 241

QY 502 GKPNISOIYETDEVRIAQLADKYTTSDGYIFDEHDIISDEGDVYVTPMHGSHWIGKDSL 561
Db 242 GKPNAIQIYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDVYVTPMTHSHWIKKDSL 301

QY 562 SDKEKVAQAAYTKEGKILPPSPADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVHEHT 621
Db 302 SEAEERAAQAAYAKEKGLTPPSTDHDSNGTEAKGAEAIYNRVKAQKVPDLRMPYNLQYT 361

QY 622 VEVKNGNLIIPHKHGHHYHNKFAWFDHTYKAPNGYTTLEDLPAITKYVVEHPDERPHSNDG 681
Db 362 VEVKNGSLIIPHYDHYHNKFEWFDGLYEAPKGYTTLEDLLATVKYVVEHPNERPHSNDG 421

QY 774 EAEKLLALLKGS 785
Db 601 EAEKLLALLKGS 612

RESULT 22
US-10-324-143-54
; Sequence 54, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-54

Query Match 45.6%; Score 1898; DB 15; Length 1126;
Best Local Similarity 64.9%; Pred. No. 1e-127;
Matches 366; Conservative 63; Mismatches 101; Indels 34; Gaps 3;

QY 266 DIDLKQLYKLPISQRHVESDGLVDFPAQITTSRTARGVAVPHGHHYHFIYEQMSELEE 325
Db 2 DIDLKQLYKLPISQRHVESDGLVDFPAQITTSRTARGVAVPHGHHYHFIYEQMSELEE 61

QY 326 RIARIIPLYRSNHWVPSRPEOPSPOPTPEPSGPQAPNLK-IDSN--SSLVSQLVR 381
Db 62 RIARIIPLYRSNHWVPSRPEOPSPOPTPEPSGPQAPNPQAPNPIDKLVKEAVR 121

QY 382 KVGEGYVFEERKIGSYVFAKDLPSFTVKNLESKLSKQESVSHLTLTAKKENVAPRDQEFYD 441
Db 122 KVGEGYVFEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSSDRFYN 181

QY 442 KAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESNKEKLVDDLLAFAPITHPERL 501
Db 182 KAYDILLARIHQDLLDNKGRQVDFEALDNLLERLKDVSDDKVLVDDIILAFAPIRHPERL 241

QY 502 GKPNISOIYETDEVRIAQLADKYTTSDGYIFDEHDIISDEGDVYVTPMHGSHWIGKDSL 561
Db 242 GKPNAIQIYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDVYVTPMTHSHWIKKDSL 301

QY 562 SDKEKVAQAAYTKEGKILPPSPADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVHEHT 621
Db 302 SEAEERAAQAAYAKEKGLTPPSTDHDSNGTEAKGAEAIYNRVKAQKVPDLRMPYNLQYT 361

QY 622 VEVKNGNLIIPHKHGHHYHNKFAWFDHTYKAPNGYTTLEDLPAITKYVVEHPDERPHSNDG 681
Db 362 VEVKNGSLIIPHYDHYHNKFEWFDGLYEAPKGYTTLEDLLATVKYVVEHPNERPHSNDG 421
```



```

; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66

Query Match 63.6%; Score 2649.5; DB 9; Length 763;
Best Local Similarity 66.8%; Pred. No. 3.1e-182;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;

QY 1 SYELGHYQRTV-KENNRVSYIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGV 59
DB 2 SYELGRHQAGQVKKESNRVSYIDGQAGQKAENLTPEVSKREGINAEQIVIKITDQGV 61
QY 60 TSHGDHYYNGKVPYDAIISELLMKDPNYKDKEDIVNEVKGYYIKVDGYVYVYVKD 119
DB 62 TSHGDHYYNGKVPYDAIISELLMKDPNYKDKSDIVNEIKGYYIKVNGKYYVYVKD 121
QY 120 AAHADNVRTKEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIID 179
DB 122 AAHADNRTKEIKRQKQSHSHNS---RADNAVAARAQGRYTTDDGYIFNASDIID 178
QY 180 TGDAYIVPGDHVHYTPKNELSAELAAAFISGRGNLSNRYRQNSDNTSRINWVP 239
DB 179 TGDAYIVPGDHVHYTPKNELSAELAAAEAYWNG-----KQGRSPSSSSYNA 227
QY 240 SVSNPGTNTNTNNSNTNSQAQSDNDIDSLLKQYKLPISQRHVESDGLVFPQAQITSR 299
DB 228 NPAQPLSENHNLTVPTTHQ--NQGENISLLRELVAKPLSERHVESDGLIFPQAQITSR 286
QY 300 TARGVAVPHGDHYHFIYQMSLEERARIIPLRYRSHNHVWPSRPEQSPQPTPEPSP 359
DB 287 TARGVAVPHGNHYHFIYEQMSLEKRIARIIPLRYRSHNHVWPSRPEQSPQPTPEPSP 346
QY 360 GPQAPNLK-IDSN---SSLVSQIVKVGEGYFEEKIGIRYVFAKDLPSRTVKNLESKL 415
DB 347 SPOFAPNPQAPSNPIDKLVKSAVRKVGQYVFEENGVSRYTPAKDLSAETAAGIDSKL 406
QY 416 SKQESVSHLTAKKENVAPRDQBYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
DB 407 AKQESLSHKGAKTDLPSDRFYKAYDILLARIHQDLDLNKGRQVDFEALDNLRLERLK 466
QY 476 DESTNKEKLVDDLLAFAPITHPERLGKNSQIETTEDVIRIAQLADKYTTSDGYIFDEH 535
DB 467 DVKSDKVLXVOLLAFAPIRHPERLGKPNQAITYTDDIQAQVAKLAGKYTTEDGYIFDPR 526
QY 536 DIISDEGDYAVTHMGCHSHWIGKDSIDKEKVAQAAYTEKGLPLSPDADKANPTGDS 595
DB 527 DIISDEGDYAVTHMTSHWIKKDSISEAERAAQAAYAKEGLTTPSTDHQDSGNTAKG 586
QY 596 AAAYNRVKGKRIPLVRPLVMYVEHVEVKNGNLIIPHDKDHYHNKFAWDDHTYKAPNG 655
DB 587 AEALYNRVKAAKVPDLRPNYLNQYVEVKNGSLIIPHVDHYHNKFEWDEGLYAPKG 646
QY 656 YTTLEDLPATIKYVVEHPDPRHNDGKNASEHVLGKQKHSDEPNKNFKADEE-----P 709
DB 647 YTTLEDLATYKYYVEHPNERPHSDNGFGNASDHVQRNKGQADTNQTEKPESEKPKTEK 706
QY 710 VEET-----PAPPEVPOVETEKVEAQKAEVLLAKVTD 743
DB 707 EETPREEKQSEKPSKPTPEEPSESEEPQVETEKVEEKLREADLLGKIQD 763
```

## RESULT 20

```

US-10-324-143-22
; Sequence 22, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-22
```

```

Query Match 52.1%; Score 2170; DB 15; Length 613;
Best Local Similarity 68.5%; Pred. No. 9.2e-148;
Matches 419; Conservative 59; Mismatches 100; Indels 34; Gaps 3;

QY 208 AEAFLSGRGNLSNRYRQNSDNTSRINWVPSVSPGCTTNTNNSNTNSQAQSDNDI 267
DB 1 AEAFLSGRGNLSNRYRQNSDNTSRINWVPSVSPGCTTNTNNSNTNSQAQSDNDI 60
QY 268 DSSLKQYKLPISQRHVESDGLVFPQAQITSRARGVAVPHGDHYHFIYQMSLEERII 327
DB 61 DSSLKQYKLPISQRHVESDGLVFPQAQITSRARGVAVPHGNHYHFIYQMSLEEKRI 120
QY 328 ARIIPLRYRSHNHVWPSRPEQSPQPTPEPSPQAPNLK-IDSN---SSLVSQIVKRV 383
DB 121 ARIIPLRYRSHNHVWPSRPEPSPQPTPEPSPQAPNPQAPSNPIDKLVKEAVRVK 180
QY 384 GEGYVFEKIGIRYVFAKDLPSRTVKNLESKLSQESVSHLTAKKENVAPRDQFVYDKA 443
DB 181 GDGYVFEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKTDLPSDRFYNKA 240
QY 444 YNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTNKEKLVDDLLAFAPITHPERLGK 503
DB 241 YDLARIHQDLDLNKGRQVDFEALDNLRLERLKDVSDDKVKLVDDLLAFAPIRHPERLGK 300
QY 504 PMSQIETTEDVIRIAQLADKYTTSDGYIFDDEHDIISDEGDYAVTHMGCHSHWIGKDSLD 563
DB 301 PNAQITYTDDIQAQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPHMTSHWIKKDSLSE 360
QY 564 KEKVAQAAYTEKGLPLSPDADKANPTGDSAAAYNRVKGKRIPLVRLPYMVEHTE 623
DB 361 AERAAQAAYAKEGLTTPSTDHQDSGNTAKGAEALYNRVKAAKVPDLRPNYLNQYVTE 420
QY 624 VKNGNLIIPHDKDHYHNKFAWDDHTYKAPNGYTTLEDLPATIKYVVEHPDPRHSDNGWG 683
DB 421 VKNGSLIIPHVDHYHNKFEWDEGLYAPKGKYTTLEDLATYKYYVEHPNERPHSDNGFG 480
QY 684 NASEHVLGKQKHSDEPNKNFKAD-----EPPVEET 713
DB 481 NASDHVQRNKGQADTNQTEKPESEKPKTEKPEETPREEKQSEKPSKPTPEEPSEES 540
QY 714 PAPPEVPOVETEKVEAQKAEVLLAKVTDSSLKANATETLAGLRNLTQIDMNNISIMA 773
DB 541 PEESEEPQVETEKVEEKLREADLLGKIQDPIIKSNAKETLTGLKNNLFGTQDNNTIMA 600
```



```

Db      366 SLOPAPNPQAPSPNIPID--EKLKVEAVRKVGDDGVFFENGVSRYIPAKDLISAETAGIDS 423
Qy      414 KLSQESVSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLER 473
Db      424 KLAQESLSHKLGAKTDLPSDDREFYFNKAYDILLARHQDLDLNDKGRQVDFEVLNLLER 483
Qy      474 LNDESTNKEKLVDDLLAFAPITHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 533
Db      484 LKDVSSDKVKLVDLILAFAPIRHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 543
Qy      534 EHDIIISDEGDAYVTPHMGSHWIKGSLSDKEKVAQAQYTKKEGILPPSPDADVKANPTG 593
Db      544 PRDITSDGDAYVTPHMGSHWIKGSLSDKEKVAQAQYTKKEGILPPSPDADVKANPTG 603
Qy      594 DSAAIYNRVKGKRIPLVRLPYMVEHTVEYKNGNLIIPHKKDHYHNKIFAWFDDHTYKAP 653
Db      604 KGAELIYNRVKAACKVPLDRMPYNLQYTVVEYKNGSLIIPHYDHYHNKIFEFWDEGLYEAP 663
Qy      654 NGYTLDELATIKYVVEHPDRPHSDNGWGNASHEVLGK-----KDHSE----- 697
Db      664 KGYSLDILLATVYVVEHPDRPHSDNGWGNASHEVLGK-----KDHSE----- 723
Qy      698 -----DPNKNFKADEPVEETPAEPEVPQVETEKVEAQLKEAEV 736
Db      724 HPSEDEKENHAGLNPSADNLYKPSDTEETEEBAEDTTDEAEIQQVENSVINAKIADAEA 783
Qy      737 LLAQVTSLSKANATETLAGLNRLNLTQIMDNNSIMAEAEKLLALLKGSNPSV 790
Db      784 LLEKVTDPISIRQNAETITGLKSSLLGTCKNNNTISAEDVSLALLKESQPAPI 837

RESULT 18
US-10-324-143-9
; Sequence 9, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-9

Query Match 66.0%; Score 2750.5; DB 15; Length 838;
Best Local Similarity 64.5%; Pred. No. 1.8e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

Qy      1 SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTDPDEVSKEGINAEQIVIKITDQGV 59
Db      21 SYELGRHQAGVKESNRVSYIDGQAGQKAENLTDPDEVSKEGINAEQIVIKITDQGV 80
Qy      60 TSHGDHYHYNGKVPYDAIISBELLMKDPNYKLKDEDIYNEKGGVVIKVDGKYVYLKD 119
Db      81 TSHGDHYHYNGKVPYDAIISBELLMKDPNYKLKDEDIYNEKGGVVIKVDGKYVYLKD 140
Qy      120 AAHADNVRTEETNRKQESHQSHREGGTPRNDCAVALARSQGRYTTDDGYIFNASDIID 179

```

```

Db      141 AAHADNIRTKKEIKRQKQESHSHNHNS---RADNAVAARAQGRYTTDDGYIFNASDIID 197
Qy      180 TGDAYIVPHGDHYHYIPKNELSAELAAAEAFLSGRGNLSNSTRYRQNSDNTSRTNWVP 239
Db      198 TGDAYIVPHGDHYHYIPKNELSAELAAAEAFWNG-----KQGRPSSSSSYNA 246
Qy      240 SVSNPGTNTNTNNSNTNSQASQNDIDSLKLQLYKLPLSORHVESDGLVDFDPAQITSR 299
Db      247 NPQVPRLSNHNLTVPFTHQ-NQGENISLLBELYAKPLSERHVESDGLVDFDPAQITSR 305
Qy      300 TARGVAVPHGDHYHYIPYSQMSLEERARIIPLRYSNHNWVDSRPEQSPQPTPPSP 358
Db      306 TARGVAVPHGNHYHFIPEQMSLEKRIARIIPLRYSNHNWVDSRPEQSPQPTPPSP 365
Qy      359 -----PQAPAPNLKIDNSLSLYQVRYKVGEGVVEEKGISRYVFAKDLPSETVKLES 413
Db      366 SLOPAPNPQAPSPNIPID--EKLKVEAVRKVGDDGVFFENGVSRYIPAKDLISAETAGIDS 423
Qy      414 KLSQESVSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLER 473
Db      424 KLAQESLSHKLGAKTDLPSDDREFYFNKAYDILLARHQDLDLNDKGRQVDFEVLNLLER 483
Qy      474 LNDESTNKEKLVDDLLAFAPITHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 533
Db      484 LKDVSSDKVKLVDLILAFAPIRHPERLGRPNQSIETEDVRIQALADKVTTSDDGYIED 543
Qy      534 EHDIIISDEGDAYVTPHMGSHWIKGSLSDKEKVAQAQYTKKEGILPPSPDADVKANPTG 593
Db      544 PRDITSDGDAYVTPHMGSHWIKGSLSDKEKVAQAQYTKKEGILPPSPDADVKANPTG 603
Qy      594 DSAAIYNRVKGKRIPLVRLPYMVEHTVEYKNGNLIIPHKKDHYHNKIFAWFDDHTYKAP 653
Db      604 KGAELIYNRVKAACKVPLDRMPYNLQYTVVEYKNGSLIIPHYDHYHNKIFEFWDEGLYEAP 663
Qy      654 NGYTLDELATIKYVVEHPDRPHSDNGWGNASHEVLGK-----KDHSE----- 697
Db      664 KGYSLDILLATVYVVEHPDRPHSDNGWGNASHEVLGK-----KDHSE----- 723
Qy      698 -----DPNKNFKADEPVEETPAEPEVPQVETEKVEAQLKEAEV 736
Db      724 HPSEDEKENHAGLNPSADNLYKPSDTEETEEBAEDTTDEAEIQQVENSVINAKIADAEA 783
Qy      737 LLAQVTSLSKANATETLAGLNRLNLTQIMDNNSIMAEAEKLLALLKGSNPSV 790
Db      784 LLEKVTDPISIRQNAETITGLKSSLLGTCKNNNTISAEDVSLALLKESQPAPI 837

RESULT 19
US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083

```

```

RESULT 16
US-10-324-143-32
; Sequence 32, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHAULAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-32

Query Match      66.0%; Score 2750.5; DB 15; Length 819;
Best Local Similarity 64.5%; Pred. No. 1.8e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTDPDEVSKREGINAEQIVIKITDQGYV 59
Db 2 SYELGRHQAGQVKESNRVSYIDGQAGQKAENLTDPDEVSKREGINAEQIVIKITDQGYV 61
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIVNEIKGGYVVKDGYVYVLKD 119
Db 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIVNEIKGGYVVKDGYVYVLKD 121
QY 120 AAHADNVRKKEINRQKQESHQSHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIED 179
Db 122 AAHADNIRTKKEIKRQKQESHSHNHS---RADNAVAARAAQGRYTTDDGYIFNASDIIED 178
QY 180 TGDAYIVPHGDHYHYIPKKNLSASELAFAFSLSGRGNLSNRTYRQNSDTSRINWVP 239
Db 179 TGDAYIVPHGDHYHYIPKKNLSASELAFAFSLSGRGNLSNRTYRQNSDTSRINWVP 227
QY 240 SVSNPQTTNTSNNSNTSQAQSQNDISLLKQLYKLPISORHVESDGLVFPDPAQITSR 299
Db 228 NPQPRLSENHNLTVTPTYHQ-NOGENISLLRELYAKPLSERHVESDGLVFPDPAQITSR 286
QY 300 TARGVAVPHGDHYHYFIPYQMSLEBRIARIIPLYRSNHNWVDPSPRQPSOPTPEPS- 358
Db 287 TARGVAVPHGNHYHYFIPYQMSLEKRIARIIPLYRSNHNWVDPSPRQPSOPTPEPS 346
QY 359 -----PGPOPAPNLKIDSNSLVQVRKVGEGYVFEKGISRYVFAKOLPSETVKNLES 413
Db 347 SIQAPNPQAPFNSPID--EKLKVEARVKVGEGYVFEENGVSRYIIPAKOLSAETAAGIDS 404
QY 414 KLSQKESVSHLTITAKENVAPRQDEYDKAYNLLTEAHKALFNKGRNDSDFQALDKLLER 473
Db 405 KLAQKESLHKLGAQKTDLPSSDRFPYKAYDILLARIHODLLDNKGRQVDFEVLNLLER 464
QY 474 INDESTNKEKLVDDLLAFAPTHPRLGKPNSSQIETEDVRIALADKYTTSDGYIFD 533
Db 465 LKDVSSDKVLDVDDLLAFAPTHPRLGKPNQAQITTDDEIQVAKLAGKYTTEDGYIFD 524
QY 534 EHDIIISDEGDAYVTPHMGSHHWIGKDSLSDEKVAQAQVYTKKGIILPPSPDADVKANPTG 593
Db 525 PRDIISDEGDAYVTPHMTSHHWIKKDSLSSEABRAAAQAYAKEKGLTPPSTDHQDSNGTEA 584

```

## RESULT 17

```

US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match      66.0%; Score 2750.5; DB 10; Length 838;
Best Local Similarity 64.5%; Pred. No. 1.8e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTDPDEVSKREGINAEQIVIKITDQGYV 59
Db 21 SYELGRHQAGQVKESNRVSYIDGQAGQKAENLTDPDEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEIVNEIKGGYVVKDGYVYVLKD 119
Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSIVNEIKGGYVVKDGYVYVLKD 140
QY 120 AAHADNVRKKEINRQKQESHQSHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIED 179
Db 141 AAHADNIRTKKEIKRQKQESHSHNHS---RADNAVAARAAQGRYTTDDGYIFNASDIIED 197
QY 180 TGDAYIVPHGDHYHYIPKKNLSASELAFAFSLSGRGNLSNRTYRQNSDTSRINWVP 239
Db 198 TGDAYIVPHGDHYHYIPKKNLSASELAFAFSLSGRGNLSNRTYRQNSDTSRINWVP 246
QY 240 SVSNPQTTNTSNNSNTSQAQSQNDISLLKQLYKLPISORHVESDGLVFPDPAQITSR 299
Db 247 NPQPRLSENHNLTVTPTYHQ-NOGENISLLRELYAKPLSERHVESDGLVFPDPAQITSR 305
QY 300 TARGVAVPHGDHYHYFIPYQMSLEBRIARIIPLYRSNHNWVDPSPRQPSOPTPEPS- 358
Db 306 TARGVAVPHGNHYHYFIPYQMSLEKRIARIIPLYRSNHNWVDPSPRQPSOPTPEPS 365
QY 359 -----PGPOPAPNLKIDSNSLVQVRKVGEGYVFEKGISRYVFAKOLPSETVKNLES 413

```

```

Query Match      66.6%; Score 2772; DB 15; Length 838;
Best Local Similarity 65.0%; Pred. No. 5.2e-191;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLVQARTV-KENNRVSYIDGKQATOKTENITPDEVSKREGINAEQIVIKITDQGYV 59
DB 21 SYELGRHQAGQVKESNRVSYIDGQAGQKAENITPDEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEVKGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEIKGYYVIKVDGKYVYLKD 140
QY 120 AAHADNVRTKEEINRQKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIID 179
DB 141 AAHADNVRTKEEIKRQKQEHSHHGGGS--NDQAVVAARAQGRYTTDDGYIFNASDIID 198
QY 180 TGDYIVPHGDHYHYIPKNEISASELAAAEFLSGRGNLSNRYRRONSNTSRTWVP 239
DB 199 TGDYIVPHGDHYHYIPKNEISASELAAAEAYWNG-----KQSRPSSSSSYNA 247
QY 240 SVSNPGTTNTNTSNNSTNSQASQNDIDSLKOLYKPLSORHVESDGLVDPQAITSR 299
DB 248 NPAQPLSENENLTVTPYHQ--NQCENISLLRELYAKPLSERHVESDGLIFDPAQITSR 306
QY 300 TARGVAVPHGDHYHYIPYSQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 359
DB 307 TARGVAVPHGNHYHPIPYEQNSELEKRIARIIPLYRSNHWVPSRPPQSPQPTPEPSP 366
QY 360 GPQAPNLK-IDSN--SSLSQLVKRVGEGVVEEKGISRYVFAKOLPSETVKNLESKL 415
DB 367 SPQAPNPQAPSNPIDDKLKEAVRKVGDGVPFENGVSRYIPAKOLSAETAAGIDSKL 426
QY 416 SKQESVSHLTAKKENVAPRDOEFYDKAYNLITLTAHKALEFKNKGRNSDFQALDKLLERLN 475
DB 427 AKQESLSHKLGAKKTDLPSSDREFYFNKAYDILLARITHQDLDDNKGQVDFEALDNLLERLK 486
QY 476 DESTNKEKLVDDLLAFAPITHPERLKGPNQOIEYTEDVRITAOALADKYTTSDGYIFDEH 535
DB 487 DVPDKVKLVDDLLAFAPIRHPERLKGPNQOITVDDDEIQVAKLAGKTTEDGYIFDPR 546
QY 536 DIISDEGDYVTPHMGHSHWGKOSLSDEKVAQAAYTKEGILPSPDADVKANPTGDS 595
DB 547 DITSDEGDYVTPHMTSHWIKKOSLSSEARAAAQAYAKEGLTPPSTDHQDSGNTGAKG 606
QY 596 AAAIYNRVKGEKRIPIPLRVLYPVVHEVTVKNGNLIIIPKDHVHNLIKFAWFDHHTYKAPNG 655
DB 607 ABAYINRVKAAKKVPLDRMPYLNQYTVVEKNGSLIIPHYDHYHNLIKFEWFDGLYEAPKG 666
QY 656 YTLEDLFATIKYVVEHPDERPHSDNGWGNASHEVLGK-----KDHSE----- 697
DB 667 YTLEDLLATVKYVVEHPDERPHSDNGFNGNASHVRKNKVDQDSKPDEKDEHDEVSEPTHP 726
QY 698 -----DPKNKFADEPVEETPAEPEVPQVETEKVEAOLKEAEVLL 738
DB 727 ESEKENHAGLNPSADNLKYPSTDTTEETEEBAEDTTDEAEIQVENSVINAKIADAEALL 786
QY 739 AKVTDSSLKANATETLAGLRNNLITLQIMDNNSIMAEAEKLLALLKGSNPSSV 790
DB 787 EKVTDPSIRQNAMEITLGLKSSLLGTGKNTTISAEVDSLLALLKESQAPI 838

RESULT 15
US-10-387-783-4
; Sequence 4, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683

```

```

; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-387-783-4

Query Match      66.6%; Score 2772; DB 15; Length 838;
Best Local Similarity 65.0%; Pred. No. 5.2e-191;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLVQARTV-KENNRVSYIDGKQATOKTENITPDEVSKREGINAEQIVIKITDQGYV 59
DB 21 SYELGRHQAGQVKESNRVSYIDGQAGQKAENITPDEVSKREGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEVKGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYYHYNGKVPYDAIISEELLMKDPNPKLDEDIVNEIKGYYVIKVDGKYVYLKD 140
QY 120 AAHADNVRTKEEINRQKQEHSHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIID 179
DB 141 AAHADNVRTKEEIKRQKQEHSHHGGGS--NDQAVVAARAQGRYTTDDGYIFNASDIID 198
QY 180 TGDYIVPHGDHYHYIPKNEISASELAAAEFLSGRGNLSNRYRRONSNTSRTWVP 239
DB 199 TGDYIVPHGDHYHYIPKNEISASELAAAEAYWNG-----KQSRPSSSSSYNA 247
QY 240 SVSNPGTTNTNTSNNSTNSQASQNDIDSLKOLYKPLSORHVESDGLVDPQAITSR 299
DB 248 NPAQPLSENENLTVTPYHQ--NQCENISLLRELYAKPLSERHVESDGLIFDPAQITSR 306
QY 300 TARGVAVPHGDHYHYIPYSQMSLEERLARIIPLYRSNHWVPSRPPQSPQPTPEPSP 359
DB 307 TARGVAVPHGNHYHPIPYEQNSELEKRIARIIPLYRSNHWVPSRPPQSPQPTPEPSP 366
QY 360 GPQAPNLK-IDSN--SSLSQLVKRVGEGVVEEKGISRYVFAKOLPSETVKNLESKL 415
DB 367 SPQAPNPQAPSNPIDDKLKEAVRKVGDGVPFENGVSRYIPAKOLSAETAAGIDSKL 426
QY 416 SKQESVSHLTAKKENVAPRDOEFYDKAYNLITLTAHKALEFKNKGRNSDFQALDKLLERLN 475
DB 427 AKQESLSHKLGAKKTDLPSSDREFYFNKAYDILLARITHQDLDDNKGQVDFEALDNLLERLK 486
QY 476 DESTNKEKLVDDLLAFAPITHPERLKGPNQOIEYTEDVRITAOALADKYTTSDGYIFDEH 535
DB 487 DVPDKVKLVDDLLAFAPIRHPERLKGPNQOITVDDDEIQVAKLAGKTTEDGYIFDPR 546
QY 536 DIISDEGDYVTPHMGHSHWGKOSLSDEKVAQAAYTKEGILPSPDADVKANPTGDS 595
DB 547 DITSDEGDYVTPHMTSHWIKKOSLSSEARAAAQAYAKEGLTPPSTDHQDSGNTGAKG 606
QY 596 AAAIYNRVKGEKRIPIPLRVLYPVVHEVTVKNGNLIIIPKDHVHNLIKFAWFDHHTYKAPNG 655
DB 607 ABAYINRVKAAKKVPLDRMPYLNQYTVVEKNGSLIIPHYDHYHNLIKFEWFDGLYEAPKG 666
QY 656 YTLEDLFATIKYVVEHPDERPHSDNGWGNASHEVLGK-----KDHSE----- 697
DB 667 YTLEDLLATVKYVVEHPDERPHSDNGFNGNASHVRKNKVDQDSKPDEKDEHDEVSEPTHP 726
QY 698 -----DPKNKFADEPVEETPAEPEVPQVETEKVEAOLKEAEVLL 738
DB 727 ESEKENHAGLNPSADNLKYPSTDTTEETEEBAEDTTDEAEIQVENSVINAKIADAEALL 786
QY 739 AKVTDSSLKANATETLAGLRNNLITLQIMDNNSIMAEAEKLLALLKGSNPSSV 790
DB 787 EKVTDPSIRQNAMEITLGLKSSLLGTGKNTTISAEVDSLLALLKESQAPI 838

```

Db 306 TARGVAVPHGHNYHFIPIYEQMSELEKRIARIPIRYRSNHWVPOSRPEPSPQPTPEPSP 365  
 QY 360 GPQAPNLKIDNSNSLVQVVKVGEVVEEKGISRYVFAKDLPSVTKVLESKLSQOE 419  
 Db 366 SPQAPSPNID--GKLVKEAVRKVGDDGVVEENGVSRYVPAKDLJSAETAAGIDSKAQOE 423  
 QY 420 SVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDEST 479  
 Db 424 SLSHKLGTKTKDLPSSDREFYNKAYDILLARIHQDLDLNGKQVDFEALDNLLEKLVDS 483  
 QY 480 NKEKLVDDLLAFAPITHPERLGPNSQIETDEVRIAQLADKYTTSDGVIYFDEHDIIS 539  
 Db 484 DKVXLVEDILAFAPIRHPERLGPNAQIITYDDIEIQVAKLAGKYTABDGYIFDPRDITS 543  
 QY 540 DEGDAYVTPHMGHSHWIKGDSLSDEKVAQAYTKKEGILPSPDADVKANPTGDSAAAI 599  
 Db 544 DEGDAYVTPHMGHSHWIKGDSLSDEKVAQAYTKKEGILPSPDADVKANPTGDSAAAI 603  
 QY 600 YNRVKGKRIPLVRLPYMVEHTVEVXNGNLIIPHKHYNHNIKFAWFDHHTYKAPNGYTL 659  
 Db 604 YNRVKAACKVPLDRMPNLYQTVVEVXNGSLIIPHVDHYNHNIKFEWDEGLYEAAPKGYTL 663  
 QY 660 DLFATIKYVVEHDPDRPHSDNGWGNASEHVLGKDHSEDPNKNPKADEE-----PVEET 713  
 Db 664 DLLATVKYVVEHDPDRPHSDNGWGNASEHVLGKDHSEDPNKNPKADEE-----PVEET 723  
 QY 714 -----PABPEVPOVETEKVEAQLKEAEVLLAKVTTDSLSKANA 750  
 Db 724 PREKPOSEKPEPSPKTEPEEPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 783  
 QY 751 TETLAGLNNLTQIMDNNSIMAEKLLALLKGS 785  
 Db 784 KETLTGLKNNLLFTQDNNTIMAEKLLALLKES 818

## RESULT 13

US-10-412-862-4  
 ; Sequence 4, Application US/10412862  
 ; Publication No. US20040052781A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-685  
 ; CURRENT APPLICATION NUMBER: US/10/412,862  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 838  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-412-862-4

Query Match 66.6%; Score 2772; DB 12; Length 838;  
 Best Local Similarity 65.0%; Pred. No. 5,2e-191;  
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;  
 QY 1 SYELGLYQARTV-KENNRYSYIDGKATQKTNLTDEVSKEGINAEQIVIKITDQGV 59  
 Db 21 STELGHQAGQAKESNRYSYIDGQAGQKAENLTDEVSKEGINAEQIVIKITDQGV 80  
 QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEKGGVVIKVDGKYVYLKD 119  
 Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYLKDEDIVNEKGGVVIKVDGKYVYLKD 140

QY 120 AAHADNVRTKEENKQKOEHSQHREGTGRNDGAVALARSGRYVTTDDGYIFENASDILED 179  
 Db 141 AAHADNVRTKEENKQKOEHSQHREGTGRNDGAVALARSGRYVTTDDGYIFENASDILED 198  
 QY 180 TGDYIVPHGDHYHYPKNELASASAAAAEAFSLGRGNLSNRTYRRQNSQNTSRTNWPV 239  
 Db 199 TGDYIVPHGDHYHYPKNELASASAAAAEAFSLGRGNLSNRTYRRQNSQNTSRTNWPV 247  
 QY 240 SVSNPGTYNTNSNNTNSQASQNDIDSLLKQLYKLPLSRHVHVESGLVDPDAQITSR 299  
 Db 248 NPAQPRLSNENLITVPTTYHQ--NOGENISSLLRELYAKPLSERHVHVESGLVDPDAQITSR 306  
 QY 300 TARGVAVPHGDHYHIFIPYSOMSELEERTARIPIRYRSNHWVPOSRPEPSPQPTPEPSP 359  
 Db 307 TARGVAVPHGDHYHIFIPYEQMSELEKRIARIPIRYRSNHWVPOSRPEPSPQPTPEPSP 366  
 QY 360 GPQAPNLK-IDSN---SSLSQVLRVKGEGYVEEKGISRYVFAKDLPSVTKVLESK 415  
 Db 367 SPQAPNPQAPSPNIDPKLVKEAVRKVGDDGVVEENGVSRYVPAKDLJSAETAAGIDSK 426  
 QY 416 SKOESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475  
 Db 427 AKQESLSHLGAKKTDLPSSDREFYNKAYDILLARIHQDLDLNGKQVDFEALDNLLEK 486  
 QY 476 DESTNKEKLVDDLLAFAPITHPERLGPNSQIETDEVRIAQLADKYTTSDGVIYFDEH 535  
 Db 487 DVPSDKVLLVDDLLAFAPITHPERLGPNAQIITYDDIEIQVAKLAGKYTTEDGYIFDPR 546  
 QY 536 DIISDEGDAYVTPHMGHSHWIKGDSLSDEKVAQAYTKKEGILPSPDADVKANPTGDS 595  
 Db 547 DITSDEGDAYVTPHMGHSHWIKGDSLSDEKVAQAYTKKEGILPSPDADVKANPTGDS 606  
 QY 596 AAALYNRVKGKRIPLVRLPYMVEHTVEVXNGNLIIPHKHYNHNIKFAWFDHHTYKAPNG 655  
 Db 607 AEALYNRVKAACKVPLDRMPNLYQTVVEVXNGSLIIPHVDHYNHNIKFEWDEGLYEAAPK 666  
 QY 656 YTLEDLFATIKYVVEHDPDRPHSDNGWGNASEHVLGK-----KDHSE----- 697  
 Db 667 YTLEDLFATIKYVVEHDPDRPHSDNGWGNASEHVLGK-----KDHSE----- 726  
 QY 698 -----DPNKNFKADEPFBETPAEPEVPOVETEKVEAQLKEAEVLL 738  
 Db 727 ESDEKENHAGLNPSADNLYKPSDTEETEEAEEDTDEAEIPQVENSVINAKIADAEALL 786  
 QY 739 AKYTDLSKANATETLAGLNNLTQIMDNNSIMAEKLLALLKGSNPSV 790  
 Db 787 EKVTDPSIRONAMETLTGLKSLLLGTKDNNTISAEVDSLLALLKESQAPAPI 838

## RESULT 14

US-10-412-850-4  
 ; Sequence 4, Application US/10412850  
 ; Publication No. US20040001836A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-686  
 ; CURRENT APPLICATION NUMBER: US/10/412,850  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 838  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-412-850-4

```

Qy 1 SYELGYQA-RTVKENRVSYIDGKQATOKTENLTDPDEVSKREGINABQIVIKITDQGYV 59
Db 21 SYELGRYQAGQDKESNRVAYIDGDAQKAENLTDPDEVSKREGINABQIVIKITDQGYV 80
Qy 60 TSHGDVHYNGKVPYDAIISSELLMKDPNYLKDEDIVNEKGYVITKDGKYYVYLKD 119
Db 81 TSHGDVHYNGKVPYDAIISSELLMKDPNYQKUSDIVNEIKGYVITKNGKYYVYLKD 140
Qy 120 AAHADNVRTKEEINROKQESQHREGGTPRNDCAVALARSQGYRTDQGYIFNASDIED 179
Db 141 AAHADNIRTKEEIKRQKQERSHNHNS--RADNAVAARAQGYRTDQGYIFNASDIED 199
Qy 180 TGDAYIVPHGDHYHYPKONELSAELAAAEAFLSGRGNLSNRTYRQNSONTSETNWVP 233
Db 198 TGDAYIVPHGDHYHYPKONELSAELAAAEAYWNG-----KQGRPSSSSSSYNA 240
Qy 240 SVSNPGTINTNTSNNSNTNSQASQSDINDSLIKQLYKPLSORHVESDGLVFDPAQITSR 299
Db 247 NPAQFRUSEHNHUTVTPTYHQ-KQGENISLLURELYAKPLSRHVESDGLIFDPAQITSR 308
Qy 300 TARGVAVPHGDHYHFIPYSQMSLEERIARIIPLYRYSNHWVPDSRPEQSPQPTPEESP 359
Db 306 TARGVAVPHGNHYHFIPYEQMSLEKRIARIIPLYRYSNHWVPDSRPEEPSQPTPEESP 366
Qy 360 GPQAPNMLKIDSNSLSVQLVRKVGEGYFEEKIGISRYFAKDLPSETVKNLESKLSQOE 419
Db 366 SPQAPASNPID--CKLYKEAVRKVGQDVPFEENGVSRYIPAKDLSAETAAGIDSKLAKOE 422
Qy 420 SVSHITLAKKENVAPROQEEFYDKAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDRST 479
Db 424 SLSHKLGTKTKDTLPESSDREFYFNKAYDILLARIHQDLDLONKGRQVDFEALDNLLERLKVSS 489
Qy 480 NKEKLVDDLAFIAPITHPERLQKPNISOIYETDEDEVRIAQALADKYVITTSQGYIFDHBHDIIS 539

```

; PRIOR FILING DATE: 1999-03-19  
 ; NUMBER OF SEQ ID NOS: 388  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 194  
 ; LENGTH: 826  
 ; TYPE: PRF  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-769-787-194

Query Match 67.1%; Score 2795; DB 10; Length 826;  
 Best Local Similarity 67.0%; Pred. No. 1.1e-192; Indels 48; Gaps 8;  
 Matches 549; Conservative 83; Mismatches 139;

Qy	1	SYELGLYQA-RVYKNNRVSYIDKQATQKTENLTPDEVSKREGINAEQIVIKITDQGV 59
Db	21	SYELGRHQAGQDKKESNRVAYIDGQAGKAENLTPDEVSKREGINAEQIVIKITDQGV 80
Qy	60	TSHGDHYHYNGKVPYDAIISEELMKDPNPKLDEIVNEVKGYYIKVDGKYVYLKD 119
Db	81	TSHGDHYHYNGKVPYDAIISEELMKDPNPKLDEIVNEVKGYYIKVDGKYVYLKD 140
Qy	120	AAHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIID 179
Db	141	AAHADNVRTKEEIKRQKQSHNH--GSGANDHAAVAAARAGRYTTDDGYIFNASDIID 198
Qy	180	TGDAYIVPHGDHYHYIPKNLSASELAFAAFSGRGNLSNRYRQNSDNTSRTNWVP 239
Db	199	TGDAYIVPHGDHYHYIPKNLSASELAFAAFSGRGNLSNRYRQNSDNTSRTNWVP 239
Qy	240	SVNPGTNTNTSNNTSNQASQSDNDISLLKQYKPLSORHVESDGLVDFDPAQITSR 299
Db	248	NPAQRLSENHNLTVTPTYHQ--NGENISLLRELAKPLSERHVESDGLVDFDPAQITSR 306
Qy	300	TARGVAVPHGDHYHYIPYQMSLEERARIIPLYRSNHWVDSRPEQSPQPTPEPSP 359
Db	307	TARGVAVPHGNHYHFIPEYQMSLEERARIIPLYRSNHWVDSRPEQSPQPTPEPSP 366
Qy	360	GPQAPNLK-IDSN---SSLVSQLVRKGVGYVPEEKIGIRYVPAKDLSETVKNLESKL 415
Db	367	SPQAPNPQAPSNPID--GKLVKAVRVKGVGYVPEENGVSRYIPAKDLSAETAAGIDSKL 426
Qy	416	SKQESVHTLTAKENAVPDRQDFYKAYNLLTEAHKALFXNKGNSDFOALDKLLERLN 475
Db	427	AKGESLHKLGAKKTDLPSSDRFYKAYNLLTEAHKALFXNKGNSDFOALDKLLERLN 486
Qy	476	DESTNKEKLVDDLLAFAPITRPERLGKNSQIETDEVRIAQLADKYTTSDGYIFDEH 535
Db	487	DVPSDKVLVDLLAFAPITRPERLGKNSQIETDEVRIAQLADKYTTSDGYIFDPR 546
Qy	536	DIISDGDAYVTPHMGSHWIGKDSLSEKVAQAAYTKEKGLPSPDADVKNPTGDS 595
Db	547	DITSDEGDAYVTPHMGSHWIGKDSLSEKVAQAAYTKEKGLPSPDADVKNPTGDS 606
Qy	596	AAAIYNRVKEKRIPLVRPYMVEHTVEKNGNLIIPHDKHYHNKIPAFDDHYKAPNG 655
Db	607	AAAIYNRVKAARKVPLDRMPYNYQYVTEVKNKGLIIPHYDHYHNKIPAFDDHYKAPNG 666
Qy	656	YTLEDLFAIKYVVEHPRPHSDNGWGNASEHVLGKDHSDPKNKFXADEB-----P 709
Db	667	YTLEDLFAIKYVVEHPRPHSDNGWGNASEHVLGKDHSDPKNKFXADEB-----P 726
Qy	710	VEET-----PABEVPQVETKEVQAQLKAEVLLAKVTTDSSL 746
Db	727	VEETPREEKQSEKPEPKPTPEPESPESEBPQVETKEVBEKLEAEADLLKIQDPII 786
Qy	747	KANATETLAGLNLLTLQIMDNNSIMAEKLLALLKGS 785
Db	787	KSNKETLTLGLKNLLFGTQDNNTIMAEKLLALLKGS 825

RESULT 10  
 US-10-412-862-10  
 ; Sequence 10, Application US/10412862

; Publication No. US20040052781A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-685  
 ; CURRENT APPLICATION NUMBER: US/10/412,862  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 10  
 ; LENGTH: 819  
 ; TYPE: PRF  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-412-862-10

Query Match 67.0%; Score 2788.5; DB 12; Length 819;  
 Best Local Similarity 66.7%; Pred. No. 3.2e-192;  
 Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;

Qy	1	SYELGLYQA-RVYKNNRVSYIDKQATQKTENLTPDEVSKREGINAEQIVIKITDQGV 59
Db	21	SYELGRYQAGQDKKESNRVAYIDGQAGKAENLTPDEVSKREGINAEQIVIKITDQGV 80
Qy	60	TSHGDHYHYNGKVPYDAIISEELMKDPNPKLDEIVNEVKGYYIKVDGKYVYLKD 119
Db	81	TSHGDHYHYNGKVPYDAIISEELMKDPNPKLDEIVNEVKGYYIKVDGKYVYLKD 140
Qy	120	AAHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIID 179
Db	141	AAHADNVRTKEEIKRQKQSHNHNS---RADNAVAAARAGRYTTDDGYIFNASDIID 197
Qy	180	TGDAYIVPHGDHYHYIPKNLSASELAFAAFSGRGNLSNRYRQNSDNTSRTNWVP 239
Db	198	TGDAYIVPHGDHYHYIPKNLSASELAFAAFSGRGNLSNRYRQNSDNTSRTNWVP 246
Qy	240	SVNPGTNTNTSNNTSNQASQSDNDISLLKQYKPLSORHVESDGLVDFDPAQITSR 299
Db	247	NPAQRLSENHNLTVTPTYHQ--NGENISLLRELAKPLSERHVESDGLVDFDPAQITSR 305
Qy	300	TARGVAVPHGDHYHYIPYQMSLEERARIIPLYRSNHWVDSRPEQSPQPTPEPSP 359
Db	306	TARGVAVPHGNHYHFIPEYQMSLEERARIIPLYRSNHWVDSRPEQSPQPTPEPSP 365
Qy	360	GPQAPNLKIDSNLSLVSQLVRKGVGYVPEEKIGIRYVPAKDLSETVKNLESKLKQ 419
Db	366	SPQAPSNPID--GKLVKAVRVKGVGYVPEENGVSRYIPAKDLSAETAAGIDSKLQ 423
Qy	420	SVSHTLTAKENAVPDRQDFYKAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDEST 479
Db	424	SLSHKLGTKKTDLPSSDRFYKAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDEST 483
Qy	480	NKEKLVDDLLAFAPITRPERLGKNSQIETDEVRIAQLADKYTTSDGYIFDEHDIIS 539
Db	484	DKVKLVDDLLAFAPITRPERLGKNSQIETDEVRIAQLADKYTTSDGYIFDEHDIIS 543
Qy	540	DEGDAYVTPHMGSHWIGKDSLSEKVAQAAYTKEKGLPSPDADVKNPTGDSAAAI 599
Db	544	DEGDAYVTPHMGSHWIGKDSLSEKVAQAAYTKEKGLPSPDADVKNPTGDSAAAI 603
Qy	600	YNRVKEKRIPLVRPYMVEHTVEKNGNLIIPHDKHYHNKIPAFDDHYKAPNGYTL 659
Db	604	YNRVKAARKVPLDRMPYNYQYVTEVKNKGLIIPHYDHYHNKIPAFDDHYKAPNGYTL 663
Qy	660	DLFAIKYVVEHPRPHSDNGWGNASEHVLGKDHSDPKNKFXADEB-----PVEET 713
Db	664	DLFAIKYVVEHPRPHSDNGWGNASEHVLGKDHSDPKNKFXADEB-----PVEET 723

```

QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180
DB 141 AHADNVRTKEEINRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNASDIIEDT 200
QY 181 GDAYIVPHGDHYYHYPKNELSAELAAAFSLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 201 GDAYIVPHGDHYYHYPKNELSAELAAAFSLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
QY 241 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 300
DB 261 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 320
QY 301 ARGVAVPHGDHYYHYPYQMSSELEERARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 360
DB 321 ARGVAVPHGNHXYHYPYEQMSELEKRIIIPLYRSNHWVDSRPEEPSPQPTPEPSPS 380
QY 361 POPAPNLK-IDSN---SSIVSQLVKRGVGYVFEKGI SRYVFAKDLPSSETVKNLESKL 416
DB 381 POPAPNPQAPSNPIDKLVKEAVKVGVDGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 440
QY 417 KQESVSHTLTAKENVAVRDQDFYKAYNLLTEAHKALFYKNGRNSDFOALDKLLERLND 476
DB 441 KQESLSHKLGAKKTLDPSSDREFYKAYDOLLARIHQDLLDNKGRQVDFEALDNLLERLKD 500
QY 477 ESTNKEKLVDDLLAFIAPITHPERLGKPNQSLEYTEDEVRIQAQADKYTTSDGYIFDDEHD 536
DB 501 VSSDKVKLVDDLLAFIAPIRHPERLGKPNQITTYTDEIQVAKLAGKYTTEDGYIFDPRD 560
QY 537 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTEKGI LPPSPDADVKANPTGDSA 596
DB 561 IISDEGDAYVTPHMTSHWIKKDSLSAEARAAQAYAKEKGLTPPSTDHQDSGNTAKGA 620
QY 597 AAIYNRVKEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWFDHDTYKAPNGY 656
DB 621 EAIYNRVKAARKVPLDRMPYNQYTVVEVKNGLIIPHYDHYHNIKFEWDFEGLYEAPKY 680
QY 657 TLEDLFATIKYVVEHPDERPHSDNGMGNASEHVLGKKHSDHPNKNFKAD----- 706
DB 681 TLEDLLATVKYVVEHPDERPHSDNGMGNASDHVQRNKGADTNQTEKSEKPKQTEKPE 740
QY 707 -----EFPVEPTAEPEVQVETEKVBAQLKEAEVLLAKVTDSSL 746
DB 741 EETPREKQSEKPEPKPTEPERPEESPESEEPQVETEKVBEKLEAEEDLLGKIQDPII 800
QY 747 KANAFETIAGLRNLTLOIMDNNSIMAEAEKLLALLKGS 785
DB 801 KSNAKETLTGLKXNLLFGTQDNNTIMAEAEKLLALLKES 839

```

RESULT 8

```

US-10-324-143-38
; Sequence 38, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

FEATURE:

```

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-38

```

```

Query Match 71.1%; Score 2961; DB 15; Length 690;
Best Local Similarity 80.7%; Pred. No. 8.7e-205;
Matches 556; Conservative 54; Mismatches 75; Indels 4; Gaps 2;

```

```

QY 1 SYELGLYQARTVKENNRVSYIDGKQATOKTENLTLPDEVSKREGINAEIVIKITDQGYVT 60
DB 2 AYELGLHQAQTVKENNRVSYIDGKQATOKTENLTLPDEVSKREGINAEIVIKITDQGYVT 61
QY 61 SHGDHYYHNGKVPYDAIISBELLMKDPNYKLDKDEIVNEKGVGVYKVDGKYYVYLKDA 120
DB 62 SHGDHYYHNGKVPYDAIISBELLMKDPNYKLDKDEIVNEKGVGVYKVDGKYYVYLKDA 121
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180
DB 122 AHADNVRTKEEINRQKQHSQHREGGTSANDGAVAFARSQGRYTTDDGYIFNASDIIEDT 181
QY 181 GDAYIVPHGDHYYHYPKNELSAELAAAFSLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 182 GDAYIVPHGDHYYHYPKNELSAELAAAFSLSGRGNLSNRTYRRQNSDNTSRTNWVPS 241
QY 241 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 300
DB 242 VSNPGTTNTNNSNTNSQASQSDNDISLLKQLYKLPLSQRHVESDGLVFPDPAQITST 301
QY 301 ARGVAVPHGDHYYHYPYQMSSELEERARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 360
DB 302 ARGVAVPHGNHXYHYPYEQMSELEKRIIIPLYRSNHWVDSRPEEPSPQPTPEPSPS 361
QY 361 POPAPNLK-IDSN---SSIVSQLVKRGVGYVFEKGI SRYVFAKDLPSSETVKNLESKL 416
DB 362 POPAPNPQAPSNPIDKLVKEAVKVGVDGYVFEENGVSRYIPAKNLSAETAAGIDSKLA 421
QY 417 KQESVSHTLTAKENVAVRDQDFYKAYNLLTEAHKALFYKNGRNSDFOALDKLLERLND 476
DB 422 KQESLSHKLGAKKTLDPSSDREFYKAYDOLLARIHQDLLDNKGRQVDFEALDNLLERLKD 481
QY 477 ESTNKEKLVDDLLAFIAPITHPERLGKPNQSLEYTEDEVRIQAQADKYTTSDGYIFDDEHD 536
DB 482 VSSDKVKLVDDLLAFIAPIRHPERLGKPNQITTYTDEIQVAKLAGKYTTEDGYIFDPRD 541
QY 537 IISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTEKGI LPPSPDADVKANPTGDSA 596
DB 542 IISDEGDAYVTPHMTSHWIKKDSLSAEARAAQAYAKEKGLTPPSTDHQDSGNTAKGA 601
QY 597 AAIYNRVKEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWFDHDTYKAPNGY 656
DB 602 EAIYNRVKAARKVPLDRMPYNQYTVVEVKNGLIIPHYDHYHNIKFEWDFEGLYEAPKY 661
QY 657 TLEDLFATIKYVVEHPDERPHSDNGMGNASEHVLGKKHSDHPNKNFKAD----- 690
DB 662 TLEDLLATVKYVVEHPDERPHSDNGMGNASEHVLGKKHSDHPNKNFKAD----- 690

```

RESULT 9

```

US-09-769-787-194
; Sequence 194, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164

```



```

QY 537 IISDEGAYVTPHMGSHWIKGSLSDKEKVAQAAYTKKGIPLPPSPADYKANKPTGDSA 596
Db 542 ITSDGDAYVTPHMGSHWIKGSLSEAEARAAQAYAKGKLTTPSTDHQSGNTEAKGA 601
QY 597 AAIYNRVKGKRIPLVRLPYMVEHTVEVKNGLIIPHKHVHNKIFAWFDDHTYKAPNGY 656
Db 602 EAIYNRVKAARKVPLDRMPYNLQVTVEVKNGLIIPHVDYVHNKIFEFWDEGLYEAPKGY 661
QY 657 TLEDLFATIKYVVEHPDHPHSDNGWGNASEHVLGKGDHSDPNKNFKAD----- 706
Db 662 TLEDLLATVKYVVEHPHSDNGWGNASEHVLGKGDHSDPNKNFKAD----- 721
QY 707 -----EPEVEETPAEPEVQVETEKVEAQLKEAEVLLAKVTSSSL 746
Db 722 EETPREKQSEKPEPKTEPEEPESESESESESESESESESESESESESESESESESE 781
QY 747 KANATETLAGLNNLTQIMDNNSIMAEAKLLALLKGS 785
Db 782 KSNAKETITGLKNNLLFGTQDNNTIMAEAKLLALLKES 820

RESULT 6
US-09-884-465A-7
; Sequence 7, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match 77.3%; Score 3218; DB 10; Length 840;
Best Local Similarity 75.2%; Pred. No. 3.3e-223;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 21 AYEGLGLHQAQVTKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYQLKQSDIVNEIKGGYVIVKVGKYYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISEELMKDPNYQLKQSDIVNEIKGGYVIVKVGKYYVYLKDA 140
QY 121 AHADNVRTKEINRQKQHSQHRGGTPRNDGVALARSQRYTDDGYIFNADIIEDT 180
Db 141 AHADNVRTKEINRQKQHSQHRGGTSANDGAVAFARSQRYTDDGYIFNADIIEDT 200
QY 181 GDVIVPHGDHYHYIPKNEISASLAAAEAFSLGRNLSNRTYRRQNSDNTSRTNMVP 240
Db 201 GDVIVPHGDHYHYIPKNEISASLAAAEAFSLGRNLSNRTYRRQNSDNTSRTNMVP 260
QY 241 VSNPGTTNTNTSNTNSQASQNDISLLKQYKPLSLQRHVESDGLVFPDPAQITSR 300
Db 261 VSNPGTTNTNTSNTNSQASQNDISLLKQYKPLSLQRHVESDGLVFPDPAQITSR 320
QY 301 ARGVAVPHGDHYHYIPYQMSSELEERTARIIPLYRSHNVWPDSPRPPSPQPTPEPS 360
Db 321 ARGVAVPHGNHYHFIPEQMSSELEKRIARIIPLYRSHNVWPDSPRPPSPQPTPEPS 380

```

```

QY 361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYVEBEKIGISRYVFAKDLPEVTQKNLESKLS 416
Db 381 POPAPNPAPSPNIDKLVKAEARKVGSGGVFFENGVSRYIPAKNLSAETAAGIDSKLA 440
QY 417 KQESVSHLTAKENVAFRDQEFYDKAYNLLTEAHKALFXNKGSRNSDFQALDKLLERLND 476
Db 441 KQESLSHLKGAKKTDLPSSDREFYNKAYDILLARIHQDLDLNDKGRQVDFEALDNLLEKLD 500
QY 477 ESTNKEKLVDDLLAFAPITHPERLKGNSQIETEDVRIAQLADKVTSDGYIFDHDH 536
Db 501 VSSDKVLVDLILAFAPIRHPERLKGKFNACITTTDDDEIQVAKLAGKVTTEGDIYFDPRD 560
QY 537 IISDEGAYVTPHMGSHWIKGSLSDKEKVAQAAYTKKGIPLPPSPADYKANKPTGDSA 596
Db 561 ITSDEGDAYVTPHMGSHWIKGSLSEAEARAAQAYAKGKLTTPSTDHQSGNTEAKGA 620
QY 597 AAIYNRVKGKRIPLVRLPYMVEHTVEVKNGLIIPHKHVHNKIFAWFDDHTYKAPNGY 656
Db 621 EAIYNRVKAARKVPLDRMPYNLQVTVEVKNGLIIPHVDYVHNKIFEFWDEGLYEAPKGY 680
QY 657 TLEDLFATIKYVVEHPDHPHSDNGWGNASEHVLGKGDHSDPNKNFKAD----- 706
Db 681 TLEDLLATVKYVVEHPHSDNGWGNASEHVLGKGDHSDPNKNFKAD----- 740
QY 707 -----EPEVEETPAEPEVQVETEKVEAQLKEAEVLLAKVTSSSL 746
Db 741 EETPREKQSEKPEPKTEPEEPESESESESESESESESESESESESESESESESESE 785
QY 747 KANATETLAGLNNLTQIMDNNSIMAEAKLLALLKGS 785
Db 801 KSNAKETITGLKNNLLFGTQDNNTIMAEAKLLALLKES 839

```

## RESULT 7

```

US-10-324-143-8
; Sequence 8, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 8
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-8

```

```

Query Match 77.3%; Score 3218; DB 15; Length 840;
Best Local Similarity 75.2%; Pred. No. 3.3e-223;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 21 AYEGLGLHQAQVTKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYQLKQSDIVNEIKGGYVIVKVGKYYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISEELMKDPNYQLKQSDIVNEIKGGYVIVKVGKYYVYLKDA 140

```





```

Db      81  SHGDHYHYNGKVPYDAIIEELLKDPNYLKDEDIVNEVKGGVYIKVDGKYVYVLKDA 140
      121  AHADNVTKBEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
Db      141  AHADNVTKBEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 200
      181  GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db      201  GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
      241  VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 300
Db      261  VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 320
      301  ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
Db      321  ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 380
      361  POPAPNLKIDSNLSVLQVRKVGEVYFEEKGISRYYFAKDLPSKVKNLESKLSKQES 420
Db      381  POPAPNLKIDSNLSVLQVRKVGEVYFEEKGISRYYFAKDLPSKVKNLESKLSKQES 440
      421  VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441  VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
      481  KEKLVDDLLAPLAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501  KEKLVDDLLAPLAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 560
      541  EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 600
Db      561  EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 620
      601  NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 660
Db      621  NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 680
      661  LFATIKYVVEHPDPRHNSDGNWGNASEHVLGKDHSDPNKPKADPEEVEETPAEPVP 720
Db      681  LFATIKYVVEHPDPRHNSDGNWGNASEHVLGKDHSDPNKPKADPEEVEETPAEPVP 740
      721  QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
Db      741  QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 800
      781  LLKGSNPFSSVSKEKIN 796
Db      801  LLKGSNPFSSVSKEKIN 816

```

```

RESULT 3
US-10-412-850-8
; Sequence 8, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819

```

```

; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-8

Query Match      100.0%; Score 4163; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 1.9e-291;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYELGLCAARTVKNENRVSVDGKQATQKTNLENLPDEVSKREGINAEQIVIKITDQGVVT 60
Db 21 SYELGLCAARTVKNENRVSVDGKQATQKTNLENLPDEVSKREGINAEQIVIKITDQGVVT 80
Qy 61 SHGDHYHYNGKVPYDAIIEELLKDPNYLKDEDIVNEVKGGVYIKVDGKYVYVLKDA 120
Db 81 SHGDHYHYNGKVPYDAIIEELLKDPNYLKDEDIVNEVKGGVYIKVDGKYVYVLKDA 140
Qy 121 AHADNVTKBEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
Db 141 AHADNVTKBEINRQKQHSQHRGGTTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 200
Qy 181 GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
Db 201 GDYIVPHGDHYHYIPKNELASASLAAAEFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
Qy 241 VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 300
Db 261 VSNPGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSQRHVESDGLVDPQAITSRT 320
Qy 301 ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 360
Db 321 ARGVAVPHGDHYHYIPYSQMSSELBERIARIIPLYRSNHWVPDSRPEQSPQPTPEPSPG 380
Qy 361 POPAPNLKIDSNLSVLQVRKVGEVYFEEKGISRYYFAKDLPSKVKNLESKLSKQES 420
Db 381 POPAPNLKIDSNLSVLQVRKVGEVYFEEKGISRYYFAKDLPSKVKNLESKLSKQES 440
Qy 421 VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 441 VSHTLTAKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
Qy 481 KEKLVDDLLAPLAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 540
Db 501 KEKLVDDLLAPLAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEHDIISD 560
Qy 541 EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 600
Db 561 EGDYAVTPHMGSHWIGKDSLSDEKVAQAAYTKEGILPPSPADVKANPTGDSAAAIY 620
Qy 601 NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 660
Db 621 NRKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFADFDDHTYKAPNGYTTLED 680
Qy 661 LFATIKYVVEHPDPRHNSDGNWGNASEHVLGKDHSDPNKPKADPEEVEETPAEPVP 720
Db 681 LFATIKYVVEHPDPRHNSDGNWGNASEHVLGKDHSDPNKPKADPEEVEETPAEPVP 740
Qy 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
Db 741 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 800
Qy 781 LLKGSNPFSSVSKEKIN 796
Db 801 LLKGSNPFSSVSKEKIN 816

RESULT 4
US-10-387-783-8
; Sequence 8, Application US/10387783
; Publication No. US2004000531A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

```

89 649 15.6 272 15 US-10-324-143-2 Sequence 2, Appli  
90 649 15.6 908 15 US-10-324-143-114 Sequence 114, App

ALIGNMENTS

RESULT 1  
US-09-765-272-56  
; Sequence 56, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,272  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 56:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 796 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:  
US-09-765-272-56

Query Match 100.0%; Score 4163; DB 9; Length 796;  
Best Local Similarity 100.0%; Pred. No. 1.8e-291;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDEVSKEGINAEQIVIKITDQGYVT 60  
DB 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDEVSKEGINAEQIVIKITDQGYVT 60  
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGGYVIKVDGKYVYLKDA 120  
DB 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGGYVIKVDGKYVYLKDA 120  
QY 121 AHADNVRTKEENRQKQEHSHQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEET 180  
DB 121 AHADNVRTKEENRQKQEHSHQREGTTPNDGAVALARSQGRYTTDDGYIFNASDIIEET 180  
QY 181 GDAYIVPHGDHYHYIPKXNLSASELAFAFLSGRNLNSRFTYRQNSDNTSRTNWVPS 240  
DB 181 GDAYIVPHGDHYHYIPKXNLSASELAFAFLSGRNLNSRFTYRQNSDNTSRTNWVPS 240  
QY 241 VSNPGTNTNTNNSNTNSQASQSDNIDSLKQLYKLPLSQRHVESDGLVDFDPAQITSR 300

DB 241 VSNPGTNTNTNNSNTNSQASQSDNIDSLKQLYKLPLSQRHVESDGLVDFDPAQITSR 300  
QY 301 ARGVAVPHGDHYHYIPYSQMSLEERIARIIPLYRNSHNWVDPDSRPEQSPQPTPEPSPG 360  
DB 301 ARGVAVPHGDHYHYIPYSQMSLEERIARIIPLYRNSHNWVDPDSRPEQSPQPTPEPSPG 360  
QY 361 POPAPNLKIDNSLSLVQLVKVGEGYVFEKSGISRYVFAKDLPSVTKNLESKLSQES 420  
DB 361 POPAPNLKIDNSLSLVQLVKVGEGYVFEKSGISRYVFAKDLPSVTKNLESKLSQES 420  
QY 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480  
DB 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480  
QY 481 KEKLVDDLAFAPITPHPERLGKPNQSIQIETDEVRIAQLADKYTTSDGYIFDDEHDIISD 540  
DB 481 KEKLVDDLAFAPITPHPERLGKPNQSIQIETDEVRIAQLADKYTTSDGYIFDDEHDIISD 540  
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYATKEKGLPSPDADVKANPTGDSAAAIY 600  
DB 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQYATKEKGLPSPDADVKANPTGDSAAAIY 600  
QY 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFAWFDHHTYKAPNGYTTLED 660  
DB 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKFAWFDHHTYKAPNGYTTLED 660  
QY 661 LFAIKYVVEHPDERPHSDNGWGNASEHVLGKKDSEDPNKNFKADEPVEETPAEPPEVP 720  
DB 661 LFAIKYVVEHPDERPHSDNGWGNASEHVLGKKDSEDPNKNFKADEPVEETPAEPPEVP 720  
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780  
DB 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAKLLA 780  
QY 781 LLKGSNPSSVSKEKIN 796  
DB 781 LLKGSNPSSVSKEKIN 796  
RESULT 2  
US-10-412-862-8  
; Sequence 8, Application US/10412862  
; Publication No. US20040052781A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-685  
; CURRENT APPLICATION NUMBER: US/10/412,862  
; CURRENT FILING DATE: 2003-04-14  
; PRIOR APPLICATION NUMBER: 09/468,656  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/113,048  
; PRIOR FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 8  
; LENGTH: 819  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-412-862-8  
Query Match 100.0%; Score 4163; DB 12; Length 819;  
Best Local Similarity 99.9%; Pred. No. 1.9e-291;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDEVSKEGINAEQIVIKITDQGYVT 60  
DB 21 SYELGLYQARTVKENNRVSYIDGKQATKTENLTDPDEVSKEGINAEQIVIKITDQGYVT 80  
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGGYVIKVDGKYVYLKDA 120

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:16:48 ; Search time 405 Seconds

(without alignments)

632.474 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYBGLGYQARTVKNRVS.....KLLALLKGNPSSVSKEKIN 756

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/prodata/2/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/2/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubppaa/PCPUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/2/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/2/pubppaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/2/pubppaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/2/pubppaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/2/pubppaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/prodata/2/pubppaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/2/pubppaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/2/pubppaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/prodata/2/pubppaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/prodata/2/pubppaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	9	US-09-765-272-56
2	4163	100.0	819	12	US-10-412-862-8
3	4163	100.0	819	15	US-10-412-850-8
4	4163	100.0	819	15	US-10-387-783-8
5	3218	77.3	821	15	US-10-324-143-19
6	3218	77.3	840	10	US-09-884-465A-7
7	3218	77.3	840	15	US-10-324-143-8
8	2961	71.1	690	15	US-10-324-143-38
9	2795	67.1	826	12	US-09-769-787-194
10	2788.5	67.0	819	12	US-10-412-862-10
11	2788.5	67.0	819	15	US-10-412-850-10
12	2788.5	67.0	819	15	US-10-387-783-10
13	2772	66.6	838	12	US-10-412-862-4
14	2772	66.6	838	15	US-10-412-850-4
15	2772	66.6	838	15	US-10-387-783-4

16	2750.5	66.0	819	15	US-10-324-143-32	Sequence 32, Appl
17	2750.5	66.0	838	10	US-09-884-465A-8	Sequence 8, Appl
18	2750.5	66.0	838	15	US-10-324-143-9	Sequence 9, Appl
19	2649.5	63.6	763	9	US-09-765-272-66	Sequence 66, Appl
20	2170	52.1	613	15	US-10-324-143-22	Sequence 22, Appl
21	1899	45.6	1126	10	US-09-884-465A-383	Sequence 383, App
22	1898	45.6	1126	15	US-10-324-143-54	Sequence 54, Appl
23	1891.5	45.4	1365	10	US-09-884-465A-382	Sequence 382, App
24	1891.5	45.4	1365	15	US-10-324-143-53	Sequence 53, Appl
25	1888	45.4	555	15	US-10-324-143-26	Sequence 26, Appl
26	1808	43.3	612	15	US-10-324-143-44	Sequence 44, Appl
27	1795.5	43.1	568	15	US-10-324-143-33	Sequence 33, Appl
28	1795.5	43.1	1139	10	US-09-884-465A-380	Sequence 380, App
29	1795.5	43.1	1139	15	US-10-324-143-51	Sequence 51, Appl
30	1795.5	43.1	1378	10	US-09-884-465A-378	Sequence 378, App
31	1795.5	43.1	1378	15	US-10-324-143-49	Sequence 49, Appl
32	1666	40.0	334	15	US-10-324-143-20	Sequence 20, Appl
33	1665	40.0	1238	10	US-09-884-465A-381	Sequence 381, App
34	1661	39.9	1238	15	US-10-324-143-52	Sequence 52, Appl
35	1656	39.8	999	10	US-09-884-465A-376	Sequence 376, App
36	1655	39.7	999	15	US-10-324-143-47	Sequence 47, Appl
37	1651	39.6	999	15	US-10-324-143-48	Sequence 48, Appl
38	1649	39.6	428	15	US-10-324-143-27	Sequence 27, Appl
39	1649	39.6	999	10	US-09-884-465A-377	Sequence 377, App
40	1614.5	38.8	473	15	US-10-324-143-36	Sequence 36, Appl
41	1591.5	38.2	1058	15	US-10-324-143-46	Sequence 46, Appl
42	1552	37.3	487	15	US-10-324-143-21	Sequence 21, Appl
43	1552	37.3	487	15	US-10-324-143-35	Sequence 35, Appl
44	1247	29.9	780	15	US-10-324-143-37	Sequence 37, Appl
45	1247	29.9	1019	15	US-10-324-143-14	Sequence 14, Appl
46	1247	29.9	1039	10	US-09-884-465A-6	Sequence 6, Appl
47	1247	29.9	1039	15	US-10-324-143-7	Sequence 7, Appl
48	1217.5	29.2	489	15	US-10-324-143-15	Sequence 15, Appl
49	1217.5	29.2	509	15	US-10-324-143-16	Sequence 16, Appl
50	1203.5	28.9	484	10	US-09-769-787-38	Sequence 38, Appl
51	1203.5	28.9	484	12	US-10-412-862-6	Sequence 6, Appl
52	1203.5	28.9	484	15	US-09-769-787-38	Sequence 38, Appl
53	1203.5	28.9	484	15	US-10-412-850-6	Sequence 6, Appl
54	1203.5	28.9	484	15	US-10-387-783-6	Sequence 6, Appl
55	1187.5	28.5	447	9	US-09-765-272-182	Sequence 182, App
56	991.5	23.8	342	15	US-10-324-143-40	Sequence 40, Appl
57	991.5	23.8	913	10	US-09-884-465A-384	Sequence 384, App
58	991.5	23.8	913	15	US-10-324-143-55	Sequence 55, Appl
59	991.5	23.8	1152	10	US-09-884-465A-379	Sequence 379, App
60	991.5	23.8	1152	15	US-10-324-143-50	Sequence 50, Appl
61	966.5	23.2	840	10	US-09-884-465A-10	Sequence 10, Appl
62	929	22.3	822	10	US-09-769-736-18	Sequence 18, Appl
63	927	22.3	793	10	US-09-252-088-15	Sequence 15, Appl
64	927	22.3	793	15	US-10-340-792-15	Sequence 15, Appl
65	925.5	22.2	294	15	US-10-324-143-39	Sequence 39, Appl
66	865	20.8	205	15	US-10-324-143-18	Sequence 18, Appl
67	853	20.5	226	15	US-10-324-143-30	Sequence 30, Appl
68	841	20.2	381	10	US-09-769-736-24	Sequence 24, Appl
69	802	19.3	203	15	US-10-324-143-31	Sequence 31, Appl
70	738	17.7	908	15	US-10-324-143-112	Sequence 112, App
71	738	17.7	908	15	US-10-324-143-113	Sequence 113, App
72	733	17.6	908	15	US-10-324-143-110	Sequence 110, App
73	733	17.6	908	15	US-10-324-143-111	Sequence 111, App
74	723.5	17.4	906	10	US-09-884-465A-369	Sequence 369, App
75	714.5	17.2	906	10	US-09-884-465A-373	Sequence 373, App
76	710.5	17.1	906	10	US-09-884-465A-371	Sequence 371, App
77	685.5	16.5	938	15	US-10-324-143-157	Sequence 157, App
78	684	16.4	902	15	US-10-324-143-90	Sequence 90, Appl
79	680.5	16.3	929	15	US-10-324-143-151	Sequence 151, App
80	675.5	16.2	900	10	US-09-884-465A-333	Sequence 333, App
81	665.5	16.0	925	15	US-10-324-143-152	Sequence 152, App
82	660	15.8	715	10	US-09-252-088-16	Sequence 16, Appl
83	660	15.8	715	15	US-10-340-792-16	Sequence 16, Appl
84	659.5	15.8	929	15	US-10-324-143-150	Sequence 150, App
85	655	15.7	917	15	US-10-324-143-156	Sequence 156, App
86	654.5	15.7	907	15	US-10-324-143-158	Sequence 158, App
87	650	15.6	906	15	US-10-324-143-3	Sequence 3, Appl
88	650	15.6	908	15	US-10-324-143-149	Sequence 149, Appl





```

; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 ACY
; CURRENT APPLICATION NUMBER: US/09/816, 703A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Krueger, N.X. and Saito, H.: A human transmembrane protein-tyrosi
; OTHER INFORMATION: ne-phosphatase, PTP zeta, is expressed in brain and has an N-term
; OTHER INFORMATION: inal receptor domain homologous to carbonic anhydrases; Proc. Nat
; OTHER INFORMATION: l. Acad. Sci. USA 89 (16), 7417-7421 (1992)
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Levy, J.B., et al.: The cloning of a receptor-type protein tyrosi
; OTHER INFORMATION: ne phosphatase expressed in the central nervous system; J. Biol.
; OTHER INFORMATION: Chem. 268 (14), 10573-10581 (1993)
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION: By similarity
; NAME/KEY: DOMAIN
; LOCATION: (25)..(1635)
; OTHER INFORMATION: Extracellular (potential)
; NAME/KEY: misc.feature
; LOCATION: (25)..(2314)
; OTHER INFORMATION: Mature chain; protein-tyrosine phosphatase zeta
; NAME/KEY: DOMAIN
; LOCATION: (34)..(302)
; OTHER INFORMATION: Carbonic-anhydrase like
; NAME/KEY: SITE
; LOCATION: (105)..(105)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (134)..(134)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (223)..(223)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (232)..(232)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: DOMAIN
; LOCATION: (312)..(406)
; OTHER INFORMATION: Fibronectin Type-III
; NAME/KEY: SITE
; LOCATION: (324)..(324)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (381)..(381)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (497)..(497)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (501)..(501)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (552)..(552)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (587)..(587)
; OTHER INFORMATION: Chondroitin sulfate (potential)

```

```

; NAME/KEY: SITE
; LOCATION: (602)..(602)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (629)..(629)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (637)..(637)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (677)..(677)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: VARSPLIC
; LOCATION: (755)..(1614)
; OTHER INFORMATION: Splicing variant; missing (in short isoform)
; NAME/KEY: BINDING
; LOCATION: (997)..(997)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1017)..(1017)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1050)..(1050)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1082)..(1082)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1122)..(1122)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1456)..(1456)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (1548)..(1548)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: BINDING
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1561)..(1561)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC ...) (potential)
; NAME/KEY: SITE
; LOCATION: (1617)..(1617)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC ...) (potential)
; NAME/KEY: TRANSMEM
; LOCATION: (1636)..(1661)
; OTHER INFORMATION: Transmembrane region; potential
; NAME/KEY: DOMAIN
; LOCATION: (1662)..(2314)
; OTHER INFORMATION: Cytoplasmic (potential)
; NAME/KEY: CONFLICT
; LOCATION: (1722)..(1728)
; OTHER INFORMATION: Missing (in ref. 2)
; NAME/KEY: DOMAIN
; LOCATION: (1744)..(1997)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: ACT_SITE
; LOCATION: (1932)..(1932)
; OTHER INFORMATION: Active site; by similarity
; NAME/KEY: DOMAIN
; LOCATION: (1998)..(2314)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: misc.feature
; LOCATION: (2222)..(2222)
; OTHER INFORMATION: Ancestral active site
; US-09-816-703A-2

```

Query Match 3.4%; Score 141.5; DB 4; Length 2314;

Best Local Similarity 19.2%; Pred. No. 0.14; Gaps 43;

Matches 169; Conservative 113; Mismatches 267; Indels 331;

Qy 51 IKITDQGYVTSHGDPHYHYNGKVPYDAIISPELLMKDPNFKLKDIEDIVNEVKGYIKVD 110

Db 212 PEYKLETKGTQPGHEGAAVREDLPVYTKPLETKGTQPGHEGAAVREBPAYTER 271  
Qy 299 RTARGVAVGHGDIHYFIPYSOMSELEERARIIPLYRSHWVDPDRPEQSPQTPPEPS 358  
Db 272 LATKGHTQPGHE-----GKATVREETLEYTEFPVATKGT-----QEPHEGEXRVEEEL 319  
Qy 359 PGOPPA-----PML-----KIDNSSLSVLVYR--KVGEVYFEE 391  
Db 320 PALEVTRNRTEITONIPYTTEEIQDPTLLKXNRKIEROQAGRTTIQYEDIYVNGNVET 379  
Qy 392 KGISRYVFAKDLSETVKNLESKQESVSHLTATKENVAPDRQEFYDKAYNLL--TE 449  
Db 380 KEVSREVA--PVNEVVKVGLVVKPTVEITNLTKVEN-----KKSITVSYNLIDTTS 431  
Qy 450 AHKALFXNKRNSDFQALDKLLERLNDSTNKEKLVDDLLAFAPL--TH--PERLGKPN 505  
Db 432 AYVS-----AKTQVPHG--DKLVKEVDIENPAKEQVTSGLDYYPYTVTKTHLYNNGEN 484  
Qy 506 SQEYTEDERVIAQLADKYTTSQGYTFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSKXE 565  
Db 485 E--ENTETSTQDQL--EYKIKIDISVELYKENDRY-----RRYL--SLEAP 530  
Qy 566 KVAQAQYTKKGGILPPSPADVKANPTGDSAAAAYNRVKGKRIPLVRLPYVVEHT----- 621  
Db 531 TDTAKYFVKV-----SDRFKEMVLPVKS-----ITENTDGTY 563  
Qy 622 -VEVKNGLIIPKDHYNHNIKFAWFDHY-----KAPNGYT-LEDLPATIK-----Y 667  
Db 564 KVTVAVDQLVEGTDGYK-----DDYTFVAKSKAEQGVYTSFKQLVTAMQNSLSGV 616  
Qy 668 YVEHPDRPHSNDGNGWGNASEHVLGKKDHS-----EDPNKNFKADE--EPVEET 713  
Db 617 YTLASDM-----TAEVSLGDKQTSVLTGAFGLSGDGTSGYAIYDLKKELFD 667  
Qy 714 PASPEVPQVETEKVEAQLKEAEVLLAKVTDSSL-----KANATETTLAQLRNNTLQI 765  
Db 668 LNCATVRDLDTKVSADSKENVAALAKAANSANNVAVBGKISGAKSVAGLVASATNTV 727  
Qy 766 MDNNSIMAEKLLALLKGSN 786  
Db 728 IENSSFTG---KLIAHQDSN 745

## RESULT 22

US-09-200-650E-5  
; Sequence 5, Application US/09200650E  
; Patent No. 6680195  
; GENERAL INFORMATION:  
; APPLICANT: Patti, Joseph M.  
; APPLICANT: Foster, Timothy J.  
; APPLICANT: Hook, Magnus A.O.  
; APPLICANT: Eikhinn, Deirdre N.  
; APPLICANT: Perkins, Samuel L.  
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus  
; FILE REFERENCE: P06283US2/BAS  
; CURRENT APPLICATION NUMBER: US/09/200,650E  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,815  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: 60/098,427  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1315  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-200-650E-5

Query Match 3.4%; Score 141.5; DB 4; Length 1315;  
Best Local Similarity 18.5%; Pred. No. 0.057;  
Matches 179; Conservative 128; Mismatches 327; Indels 331; Gaps 43;

Qy 12 VKENNRVSYI-----DGKQATOKTENLTDPDESVKREGINAEOIIVIKITDOGYVTSBGHDY 66  
Db 160 LQENKSVVAVQPTNEENKKVDKTESTT-----LVKSDALKSNDETLVDNNSN 210  
Qy 67 HYYN-----GKVPYDAIISEELM-----KDPNYK--- 91  
Db 211 NENNADIILPKSTAPKRLNTRMRIAIAVQPSSTSEAKNVNDLITSNTTLTVDDADKNKIVP 270  
Qy 92 -----LKDEDIYNE--VKGG--YVIKVDGKYVYVLDAAHADNVRKKEINRQKQESQH 142  
Db 271 AQDYLKSKQITVDDKVKSGDYFTIKYSPTVQY---GLNPEDIKNIGDK----- 318  
Qy 143 REGGTNRNDCAVALARSQGRYTTDDGYIFNASDIIEDTGDAYI-----VPHGDHYH--- 194  
Db 319 -----DENGETIATAKHD-----TANLLITVTFDYVDRFNSVQMGINSIYMD 363  
Qy 195 -----IPKNELASELAAAEAFLSGRNLSNSYRQ-----NSDNTGRTNWVPSVS- 242  
Db 364 ADTIPVSKNDVE-----FNVTIGNTTTKTANIQPDYVNVNEKNSIGSAFTETVSH 414  
Qy 243 -----NPG-----TNTNTSNNSTNSQ-----ASQNDIDSILK-----QLYKPLSQ 281  
Db 415 VGNKENPGYKQTIYVNPSENSLTNAKLKVQAVHSSYPNNIGQINKDVTDIKIYQVP--K 472  
Qy 282 RHVESDGLVDPQAQITSRARGV--AVPHGDHYHFIPYSOMSELEERARIIPLYRSHNH 340  
Db 473 GYTLANKGYDVTNTELDTVNOYLQKITYGDNNSAV--IDFGNADSAVVMVNTKFOYTN- 529  
Qy 341 VPSRPEQSPQPTPEPSPGPAPNLKIDSNSL-----VSQLVKRVGEGYVFE- 390  
Db 530 -----SESPTLVQMATLSSTGNKSVSTGNALGFTNNQSGGAGQEVYKIG--NYWED 579  
Qy 391 -----EKGISRYV-----FAKDLP-----SETVK 409  
Db 580 TNKNGVQELGKRGVNGVNTVTFDNNNTNKVGEAVTKEDGSLYLPNPNNGRYRVEFSNLPK 639  
Qy 410 NLESKLSKQ-----ESVSHLTAKENNAVPRDQEFYDKAYNLLTEAHKALFXNKRGN 461  
Db 640 GYEVTFSKQGNNEELDSNGLSSVITNGKDNLSADLGIYKPKYKNGLDYVWEDTNKNGIQD 699  
Qy 462 SDFQALDKLLERLNDSTNKEKLV---DLLAFAPLITHPERLKGPNKNSQIETEDE--- 514  
Db 700 QDEKGISGVTVTLKDENGVNLKTVTTDADGKYF-----TDLDNGYVVEFTTPEGYT 752  
Qy 515 -VRIAQLADKYTTSQGY---IFEHDIISDEGDAYVTPHMGHSHWIGKDSLSKXVAA 569  
Db 753 PTTVTSGSDIEKDSNGLTTTGTGVINGADNMVTLDSGFYKTPKYNLNGYVWEDTNKDKQDST 812  
Qy 570 QAYTKEKGLPPSPDADVKANPTGDSAAAAYNRVKGKRIPLVRLPYVWVHTVEVKNGNL 629  
Db 813 -----EKGI-----SG-----VTVLKNNGEV 830  
Qy 630 IIPHK--DHYHNKFAWFDHYK-----APNGYTLDELFAIKYVVEHPDRPHSNDGWN 684  
Db 831 LQTKTKDKGYQFTGLENGTYKVEPETSPTQVGS-----GTDEGIDSN---GT 880  
Qy 685 ASEHVLGKKDHS-----EDPNKNFKADEEPVEETPAEPVQVETEK 726  
Db 881 STTGVIKDKNDVTIDSGFYKPTNGLDYVWEDTNKNGVQDK----- 922  
Qy 727 VEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNTLQIIMDNNSIMAEKLLALLKGSN 786  
Db 923 -EKGISGVTVTLKDENDKVLKTVTTD-----ENKGYQFTDLNNGTYKVE--FETPSGYT 973  
Qy 787 PSSVS 791  
Db 974 PTVSV 978

## RESULT 23

US-09-816-703A-2  
; Sequence 2, Application US/09816703A





QY 497 -HPERLGKPNs-----OIEVTEDE-----VR--IAQL 520  
Db 694 MSSQKLGNGISSELIELQKMKESYRQVQLVQLRSLYNLQTHESQKELMVGVRNDIDAL 753  
QY 521 ADKYTTS--DGYIPDEHDIISDEGDYVTPMHGSHWIGKDSLSDEKVAQAAYTKKGI 578  
Db 754 VKTCTSLNDADII--LSDYISDQSKFESKQODLIANIGKIVSNFLQESLYTK-----808  
QY 579 LPSPDADVANKPTGDSAAAI--YNRVKGKRIPLVR-----LPMVHEHTVEVK 625  
Db 809 -----ADILHSLNDTNSIRKANEIMNRSSEFLRNASQAQAEIVGANKERIOKTV--860  
QY 626 NGMLIIIPKDH-VHNIKFAFDDHTYKAPNG-----VTLEDLFATIKYVVEHPDERP 676  
Db 861 NGSQKLGNGISSELIELQKMKESYRQVQLVQLRSLYNLQTHESQKELMVGVRNDIDAL 911  
QY 677 HSDNGWGNASEHVLGKDHSEDPNKNFKADEEPVEETPABPEVQVTEKVEAQLKEAEV 736  
Db 912 HSEDNTEKEHQQLLDLLESVLGNNDL-----IDSIKTPHTELQ-----950  
QY 737 LLAQVTDSSLK-----ANATETLAGLN-----NLTLOIMDNNSIMAAE-----EKLLA 780  
Db 951 ---KITDHLVKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVKLETTGDTSPSKRLPA 1007  
QY 781 LLKGSNPSSVSKEKIN 796  
Db 1008 TPSWTRDSSLIKETTN 1023

RESULT 19  
US-10-270-085-6  
; Sequence 6, Application US/10270085  
; Patent No. 6627408  
; GENERAL INFORMATION:  
; APPLICANT: Nislow, Corey  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Beraud, Christophe  
; TITLE OF INVENTION: Antifungal Assay  
; FILE REFERENCE: 1015  
; CURRENT APPLICATION NUMBER: US/10/270,085  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: US/03/723,820  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 05/541,782  
; PRIOR FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1073  
; TYPE: PR1  
; ORGANISM: S.pombe  
US-10-270-085-6

Query Match 3.5%; Score 145.5; DB 4; Length 1073;  
Best Local Similarity 18.6%; Pred. No. 0.02;  
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;  
QY 23 GKQATQKTNLTDFEVSKEGINAE-----QIVIKI--TDQYVTSHGHDHYHYN 70  
Db 147 GQTGKTGYTMSGD-LSDSDGILSEAGLIPRALYQLFSSLDNSNQYAVK-CSYBELN 204  
QY 71 GKPYDAIIEELLMPKNYKLDDED-----IVNEKGVGVYKVDGKYVYVLDAAHADN 125  
Db 205 EEI-RDLIVSEEL--RKPARVFEDTSRGNVITGIEESY-INKAGDGLRLREGSHRRQ 260  
QY 126 VRTKEEINRQKQHS-----QHRGGTTPRNDGAVALARSGRYTTDDGYIFNASDIIED 180  
Db 261 VAATKCNLSSRSHSIPTILHRKVSSGWTDEINSLTINN-----NSDDLAR--307  
QY 181 GDVIVPHGDHYHYIPKNLSASELAABAFLSGRGNLSNRT-----223  
Db 308 -----ASKLHMVDLAGSENI--GRSGAENKRAETQMINQSLLTLGRVI 349

QY 224 -----YR-----RONSNDTSRTNWVPSVSNPCTTNTINTSNNSNTNSQASQS 264  
Db 350 NALVEKAHHPIPYRESKULTRILQDSLGGTKTKSMIVTVS---STNTLEETISTLEYAARA 406  
QY 265 NDIDSLKQLYKPLSQRHVESDGLVFPDAQI-----TSRTARGVAVPHGDHYHFIPYSQ 319  
Db 407 K---SIRNKQNNQLVFRKVLIKDLVLDIERLKNLNAKKNKGVYLAESTYKELMDRVQ 463  
QY 320 MSEL-----ERIRIIPLYRSNHWVPSRPEQSPQPTPEPSPGPQAPNLKI-----369  
Db 464 NKDLCQEQARKLEVLNLNKS-----SREQLQYVSKSNQEHKKEVEALQOLVNSSTE 517  
QY 370 -----DNSISVLQSVKVGEGYVFEKGISRYVFAKDLPESETVKNLESKLSKQESVHT 424  
Db 518 LESVKSNEKLNELVLEIEKRYKYEAKITTTVAIDLSQYYRESKEYIASLEKLDRT 577  
QY 425 LTAKKENVAPRQDFYDKAYNLL-----TEAH 451  
Db 578 ERNKEN-----ENNFWNLKENLLTMLRSFHGSGFTDETNGYFTLLDNDFNAMESLELNTHSN 633  
QY 452 KALFXNKGKNSDFQALDKLLER-----LNDESTNKEKLVDDLLAFLAPIT 496  
Db 634 QLLISMTKITTEHFQSLDEALQASRSCAVPNSSLDLIVSELKSKNSLLDALEHSLQDIS 693  
QY 497 -HPERLGKPNs-----OIEVTEDE-----VR--IAQL 520  
Db 694 MSSQKLGNGISSELIELQKMKESYRQVQLVQLRSLYNLQTHESQKELMVGVRNDIDAL 753  
QY 521 ADKYTTS--DGYIPDEHDIISDEGDYVTPMHGSHWIGKDSLSDEKVAQAAYTKKGI 578  
Db 754 VKTCTSLNDADII--LSDYISDQSKFESKQODLIANIGKIVSNFLQESLYTK-----808  
QY 579 LPSPDADVANKPTGDSAAAI--YNRVKGKRIPLVR-----LPMVHEHTVEVK 625  
Db 809 -----ADILHSLNDTNSIRKANEIMNRSSEFLRNASQAQAEIVGANKERIOKTV--860  
QY 626 NGMLIIIPKDH-VHNIKFAFDDHTYKAPNG-----VTLEDLFATIKYVVEHPDERP 676  
Db 861 NGSQKLGNGISSELIELQKMKESYRQVQLVQLRSLYNLQTHESQKELMVGVRNDIDAL 911  
QY 677 HSDNGWGNASEHVLGKDHSEDPNKNFKADEEPVEETPABPEVQVTEKVEAQLKEAEV 736  
Db 912 HSEDNTEKEHQQLLDLLESVLGNNDL-----IDSIKTPHTELQ-----950  
QY 737 LLAQVTDSSLK-----ANATETLAGLN-----NLTLOIMDNNSIMAAE-----EKLLA 780  
Db 951 ---KITDHLVKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVKLETTGDTSPSKRLPA 1007  
QY 781 LLKGSNPSSVSKEKIN 796  
Db 1008 TPSWTRDSSLIKETTN 1023

## RESULT 20

US-08-961-083-118  
; Sequence 118, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2

```
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-541-782-6

Query Match      3.5%; Score 145.5; DB 3; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATQKNTLTPDEVSKREGINAE-----QIVIKI--TDQGVVTSBGHDHYHYN 70
Db 147 QGTGCTKYTWGSD-LSDSGILSEGAGLIPRALYQLFSSLDNSQEVAVK-CSYVELYN 204
QY 71 GKVPYDAIISBELMKDPNYKLBDE-----IVNEVGKGYVIVKDGKYYVYLKDAAHADN 125
Db 205 BEI-RDLLVSEEL--RKPARVFEDTSRRGNVITGIEESY-IKNAGDGLRLREGSHRRQ 260
QY 126 VRTKEEINRQKQHS-----OHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db 261 VAATKCNDSLRSRSHIFITILHRKVSSGTMDETNSLTINN-----NSDDLRL-- 307
QY 181 GDAYIVPHGDHYHYIPKNELASELAAAEFLSGRGNLSNRT----- 223
Db 308 -----ASKLHMVDLAGSENI--GRSGAENKRARETGMINQSLTLGRVI 349
QY 224 -----YR-----RQNSDNTSRTNWVPSVSNPGTTNTNTNNSNTNSQASQS 264
Db 350 NALVEKAHHIIPYRESKLTLLQDSLGGKTKTSMIVTVS---STNTNLEETISTLEYAARA 406
QY 265 NDIDSLAQLYKLPLSQRHVESDGLVFPDPAQI-----TSRTARGVAVPHGDHYHFIPYSQ 319
Db 407 K---SIRNKPQNNQLVFRKVLIKDLVLDIERKNDLNATRKNGVYLAESYIKELMDRVQ 463
QY 320 MSEL-----EERIARIIPLYRSHNHVWPDSPRPPSPQTPPSPGPQAPNLKI----- 369
Db 464 NKDLLCQOARKLEVLIDLNVKS-----SRQLOVYVSKNQEHKEVEALQQLVNSSTE 517
QY 370 -----DSNSSILVSOQVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQESVSH 424
Db 518 LESVKSNEKLNELVLEIEKRKKYETNEAKITTVATDLSQYVRESKEYIASLYEKLDR 577
QY 425 LTAKKENVAPRQDQFYDKAYNLL----- 451
Db 578 ERNNKEN-----ENNFNWLKFNLLTLMRSFHGSDFTDEINGYFTLLDNDFNASEELLNTHSN 633
QY 452 KALFXNKGNSDFQALDKLLER-----LNDESTNKEKLVDLLAFAPIT 496
Db 634 QLLISMTKITEHFQSLDEALQASRSCAVPNSSLDLIVSELKSKNSLLDALEHSLQDIS 693
QY 497 HPERLGKPN-----QIEYTEDE-----VR--IAQL 520
Db 694 MSSQKLGNGISSELTLOKMKESYRQLVQELRSLYNLQHTHEESQKELMYGVNRNDIDAL 753
QY 521 ADKYTTS--DGYIPDEHDIISDEGDAYVTPHMGHSHWIGKOSLSDKEXVAAQAYTKGI 578
Db 754 VKTCTISLANDAII--LSYIISQKSKFESKQODLIANIGKIVSNFLQONESLYTK----- 808
QY 579 LPPSPDADVKANPTGDSAAAI--YNRVKGKRIPLVR-----LPYVVEHTVBRVK 625
Db 809 -----ADILSHLNDTNSIRKANEIMNRSERFLRNAASQAEIVGANKERIQKTVB-- 860
QY 626 NGNLIIPKDH--YHNIKFAMFDHDTYKAPNG-----YTLEDLPATIKYVVEHPDRP 676
Db 861 NGSQLLDSKSAIHSNRSMDYDHCIALAESQKGVNLEVQTLDRLLQKVK-----E 911
QY 677 HSDGCMGNASEHVLGCKHSDPKNKFKADEBPVEETPAEPVPQVETKEVQAQLKEABV 736
Db 912 HSEDNTEKHQOOLLDLLESVLGNNDNL-----IDSIKTPHTELQ----- 950
QY 737 LIAKVTDSSLK-----ANATETLAGLRN-----NLTLQIMDNNSIMAE-----EKLLA 780
```

```
Db 951 ---KITDVLKGTTSANHTNELLGLGDESLCNLETTIEDTSLVKLETTGTPSKRELPA 1007
QY 781 LLKGSNPFSSVSKKIN 796
Db 1008 TPSWTRDSSLIKETT 1023

RESULT 18
US-09-723-820-6
; Sequence 6, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-723-820-6

Query Match      3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATQKNTLTPDEVSKREGINAE-----QIVIKI--TDQGVVTSBGHDHYHYN 70
Db 147 QGTGCTKYTWGSD-LSDSGILSEGAGLIPRALYQLFSSLDNSQEVAVK-CSYVELYN 204
QY 71 GKVPYDAIISBELMKDPNYKLBDE-----IVNEVGKGYVIVKDGKYYVYLKDAAHADN 125
Db 205 BEI-RDLLVSEEL--RKPARVFEDTSRRGNVITGIEESY-IKNAGDGLRLREGSHRRQ 260
QY 126 VRTKEEINRQKQHS-----OHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db 261 VAATKCNDSLRSRSHIFITILHRKVSSGTMDETNSLTINN-----NSDDLRL-- 307
QY 181 GDAYIVPHGDHYHYIPKNELASELAAAEFLSGRGNLSNRT----- 223
Db 308 -----ASKLHMVDLAGSENI--GRSGAENKRARETGMINQSLTLGRVI 349
QY 224 -----YR-----RQNSDNTSRTNWVPSVSNPGTTNTNTNNSNTNSQASQS 264
Db 350 NALVEKAHHIIPYRESKLTLLQDSLGGKTKTSMIVTVS---STNTNLEETISTLEYAARA 406
QY 265 NDIDSLAQLYKLPLSQRHVESDGLVFPDPAQI-----TSRTARGVAVPHGDHYHFIPYSQ 319
Db 407 K---SIRNKPQNNQLVFRKVLIKDLVLDIERKNDLNATRKNGVYLAESYIKELMDRVQ 463
QY 320 MSEL-----EERIARIIPLYRSHNHVWPDSPRPPSPQTPPSPGPQAPNLKI----- 369
Db 464 NKDLLCQOARKLEVLIDLNVKS-----SRQLOVYVSKNQEHKEVEALQQLVNSSTE 517
QY 370 -----DSNSSILVSOQVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQESVSH 424
Db 518 LESVKSNEKLNELVLEIEKRKKYETNEAKITTVATDLSQYVRESKEYIASLYEKLDR 577
QY 425 LTAKKENVAPRQDQFYDKAYNLL----- 451
Db 578 ERNNKEN-----ENNFNWLKFNLLTLMRSFHGSDFTDEINGYFTLLDNDFNASEELLNTHSN 633
QY 452 KALFXNKGNSDFQALDKLLER-----LNDESTNKEKLVDLLAFAPIT 496
Db 634 QLLISMTKITEHFQSLDEALQASRSCAVPNSSLDLIVSELKSKNSLLDALEHSLQDIS 693
QY 497 HPERLGKPN-----QIEYTEDE-----VR--IAQL 520
Db 694 MSSQKLGNGISSELTLOKMKESYRQLVQELRSLYNLQHTHEESQKELMYGVNRNDIDAL 753
QY 521 ADKYTTS--DGYIPDEHDIISDEGDAYVTPHMGHSHWIGKOSLSDKEXVAAQAYTKGI 578
Db 754 VKTCTISLANDAII--LSYIISQKSKFESKQODLIANIGKIVSNFLQONESLYTK----- 808
QY 579 LPPSPDADVKANPTGDSAAAI--YNRVKGKRIPLVR-----LPYVVEHTVBRVK 625
Db 809 -----ADILSHLNDTNSIRKANEIMNRSERFLRNAASQAEIVGANKERIQKTVB-- 860
QY 626 NGNLIIPKDH--YHNIKFAMFDHDTYKAPNG-----YTLEDLPATIKYVVEHPDRP 676
Db 861 NGSQLLDSKSAIHSNRSMDYDHCIALAESQKGVNLEVQTLDRLLQKVK-----E 911
QY 677 HSDGCMGNASEHVLGCKHSDPKNKFKADEBPVEETPAEPVPQVETKEVQAQLKEABV 736
Db 912 HSEDNTEKHQOOLLDLLESVLGNNDNL-----IDSIKTPHTELQ----- 950
QY 737 LIAKVTDSSLK-----ANATETLAGLRN-----NLTLQIMDNNSIMAE-----EKLLA 780
```





RESULT 12  
 US-09-134-001C-5080  
 ; Sequence 5080, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynh Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134, 001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064, 964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055, 779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5080  
 ; LENGTH: 3696  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5080

Query Match 3.9%; Score 163; DB 4; Length 3696;  
 Best Local Similarity 18.5%; Pred. No. 0.0063;  
 Matches 183; Conservative 141; Mismatches 348; Indels 316; Gaps 42;  
 QY 15 NNRVSYI-DGKQATQKTENLTPEVSKREGINAEQIVIKITD-----QGYVTSBGDHYH 67  
 Db 1299 NNGIDGIVSGROSINA---ITPDTSIKNAK--DIDIKAADKKIKIQRINDATDEEIQ 1353  
 QY 68 YNGKVPYDAIISEELMKDPNYKLKDEDIVNEVKGQYVTKVD-----110  
 Db 1354 EANKIE-----EAKTEAKDNIQRNSTROQVNEAKTINGINKIENITPATVTKSEARQAVQ 1408  
 QY 111 ---GKYVYLKDAHADNRTKEINR-----QKQHSQHREG--TPRNDGAVALA 157  
 Db 1409 NKANEQINHIQTPDATNEEKQBAIRNVSAELARVQAQINAEHTTQGVTKIKDITSL 1468  
 QY 158 RSQGRVTTDDGYIFNASDIIEDTGDVIVPHGDHYHYPKNELSASELAABAFLSGRGN 217  
 Db 1469 R-----INQVVEKE-----SARNA 1483  
 QY 218 LNSRSTYRRQ---NSDNTS-----RTNWVPSVSNFTGTWN--TSNNSNTNSQASQSDN 266  
 Db 1484 IEQATQTOQTFINNNDNATDEEKEVANLVIAATKQSLDNINSLSNNVDENAKVAGINE 1543  
 QY 267 IDSLKQLYKLPUSQRHVESDGLVFPDA-QITRTARGVAVPHGDHYHFTPYQSMGELEE 325  
 Db 1544 IANVL-----PATAVSKAKKIDIDQKLAQQINQIQTHQTATTBE 1582  
 QY 326 RIARIIPLVRSNH-----WVPDSRPEQSPQPTP---355  
 Db 1583 KEAAIQLANQKNEARTATQNEHSNNGVAAQKNGIHEILVMPDAAHKSDAQSDINKY 1642  
 QY 356 -EPPSGQPAPNLKIDNSLSVSQL--VRKVGEGYVFEKIGSYVFAKDLPSFTVKNLE 412  
 Db 1643 NEQSNTINTTPDATDEKQKALDKLIAKAGYKVDQATNQVQSDAKTEADITITNIQ 1702  
 QY 413 SKLSKQESYSHLTAKKE-----NVAPRDOE-----FVDKAYNLL-----447  
 Db 1703 ANVAKPSARVELDSKFEDLKRQINATPNATEEKKQDAIQLRNGKRDVKNLQNRDRDN 1762  
 QY 448 -TEAHK-----ALFXNKGRNSDFQALDKLERL-----NDESNKEKLVDDLL 489  
 Db 1763 EYEQHKNIGLOELETHANPTKSD--ALQELQTKFISQTELLNNKDATNEK---DEA 1817  
 QY 490 AFLAPITHPERL-----GKPNQIETVEDE--VRIACL-----ADKYTTSQGYIFDEH 535  
 Db 1818 KRLLISKNKTTINQAQTNQVNDNAKNGMNEIATIIPTATIKTDAKTAIDKAEQV 1877  
 QY 536 DIISDEGDVAVTPHMGHSHWIGKDSLSDKEKVAQAQYATKEKGLPSPADVKNATGDS 595  
 Db 1878 TIINGNDA-----TDEKAEARKLVERAKT-----EAKSNITNSD 1913

QY 596 AAALYNRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKEAFEDDHTYKAP-- 653  
 Db 1914 TE---REVNGAKTNGLEKINNIOPTQTKNAQEINDKAQEQLIQTNPPTDATEEKEQE 1970  
 QY 654 -----NGYTLDEL-----FATIKYVVEHPDERP---HSNDGWNAS 686  
 Db 1971 ATRNVNAGLAQAQINNAHSTQEVNESKTSIATIKSVQPNVVKKPTAINSLTQAEANQ 2030  
 QY 687 EHVLGKKKHSEDPKNF-----KADK---EPVEETPAEPEVQVETEKVEA-----Q 730  
 Db 2031 KTLIGNDGNATDDEKAAKQLVTKLNEQIQIKIHESQTDQNDQVNVKQAQAITAILINANA 2090  
 QY 731 LKEAEVL-----LAKVTDSSLKAN--AT-----ETIAGLNNL-----TLOIMD 767  
 Db 2091 HKQDAINIILTNLAESKKSDIRANQDATTBEKTAIQSIDDTLQAARNNINGANTNALVD 2150  
 QY 768 NNSIMAAEAKLLALLKGNPSSVSKEKI 795  
 Db 2151 EN-LEDGKQKQLQRIVLSTQTKTQAKADI 2177  
 RESULT 13  
 US-09-134-001C-3159  
 ; Sequence 3159, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynh Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134, 001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064, 964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055, 779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3159  
 ; LENGTH: 10182  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3159

Query Match 3.8%; Score 159; DB 4; Length 10182;  
 Best Local Similarity 18.8%; Pred. No. 0.064;  
 Matches 189; Conservative 126; Mismatches 368; Indels 320; Gaps 43;  
 QY 13 KENRVSYIDGKQATQKTENLTPEVSKREGIN-----AEQIVIKITDQGYVTS 62  
 Db 5419 QKSGEESLVNGSNTRSEVE-----EHLNEAKSLNNAKQLRDKVAEKTNVK-QSSDIYNS 5473  
 QY 63 GDHYHYNGKV-PYDAIISE-----ELLMK-----DPNYKLKDEDIVNEVKGQYVI 107  
 Db 5474 TEHQGYDQALQAEANIINEIGNPTLNKSEIEQKQLQITDAQNALQSHLLEAKNNAIT 5533  
 QY 108 KVDGKYVYVYKDAAH---ADNVRTKEEINRQKQHSQHREGGTPRND-----GAVALARS 159  
 Db 5534 GIN--KLTAINDAQORQAIENVOQOTIPAVNQOQLTDREINTAMQALRDKVQQQNNVHQ 5591  
 QY 160 QGRYTTDDGYIFNASDIIEDTGDVIVPHGDHYHYPKNEEL--SASELAABAFLSGRGN 217  
 Db 5592 QSNYFNEDQPKHYNDSVQAGOTIIDKLQD--PINNKIEIQAINQINTOTATLSENGK 5649  
 QY 218 L-----SNRTHYRQNSDTSRTNWVPSVSNPQTTNTNNTSNNTNSQASQSDIDSLK 272  
 Db 5650 LHTDQESTNRQIEGLSSLTQAQINAEKOLVNAQKTRIDVA-----QKLAABKEINSAMS 5703  
 QY 273 QLYKPLSKQHSVSDGLVFPDAQITRTARGVAVPHGDHYHFTPYQSMSELEERARIIP 332  
 Db 5704 NL-----RDGI-----QNKEDIKRSSAYINADPTKVYATYQALQNAENIINATP 5747  
 QY 333 LRYRSNHWVDSRPEQSPQPTPEPSGPGPAPNLKIDNSLSVLSQVLRKVGEYVFEK 392

```
QY 244 PGTTNTNTSNTNSQASQNDIDSLKQYKYLPLSORHVESGLVDFPAQITSRARG 303
Db 225 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 VAVPHGDHYHPIFYPSOMSELEERTARIIPILRYRSHNHVWPDSPQSPQPTPEPSGPQP 363
Db 281 VAIIPHGDHYHPIFYPSKLSALEEKLARVP----- 309
QY 364 APNLKIDNSLSVLVRKVGEGYVPEEKGISRYVFAKDLPSFTVKNLESKLSQESVSH 423
Db 310 -----ISGTGVSTNAK-----PNEVVSLSGLSSNPSS--- 339
QY 424 TLTAKKENVAPRQDEFYDKAYNLTAEHAKALFXNKGNSDFQALDKLLERLNDESTNKEK 483
Db 340 -LTTSKELSSASDGYIFNPK-DIVEETATAYIVRHG--DHPHYTPK-----SNQIGQPT 389
QY 484 LVDLLAFAPITHPERLKGKNSQIEVTEDEVRIAQLADKYTTSDGYIFDEHDIISDEGD 543
Db 390 LPNNSLATPSP-SUPINFGTSHEKHE-----EDGYGFDFANRIIADES 431
QY 544 AYVTPHMGHSHWICK 558
Db 432 GFVMSHGDNHYFEK 446

RESULT 11
US-08-956-171E-5235
; Sequence 5235, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hymann
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248PL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 5235:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 886 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5235:
US-08-956-171E-5235

Query Match
Best Local Similarity 4.1%; Score 169; DB 4; Length 886;
Matches 179; Conservative 138; Mismatches 343; Indels 246; Gaps 46;

QY 21 IDGQATQKTENI-TPDEVSKREGINAEQI-VIKITDQGVVTHGHDHYHYNKVPYDAI 78
Db 46 ITSQHOQAAAEINTSDKISENNNATTTQPPKDTNQTQFATQP-----ANTAKNYPA- 99
QY 79 ISELL-----MKDNYKLDKEDI VNEVKG VVI---KVDGKYVYV-LKDAHAADNVETKE 130
Db 100 -ADESLKDAIKPALENKEHDIGPREQVNFQLLDKNNETQYHFFSKDPADVITYTKKA 158
QY 131 EI-----NRQK-----QEHSQHREGGTPRNDGAVALARSOQ 161
Db 159 EVELDINTASTWKKFEVYENNQKLPVRLVSYPVPEDHAVIR---FPVSDGTQEL-KIVS 214
QY 162 RYTTDDGY-----IFNASDIIE-DTGDAYIVPHGDHYHYIPKNELASASELA 206
Db 215 STQIDGDEETNYDTKLVPAKPIYNDPSLVKSDTNDAVVT--NDQSSVASNOTNTSN 272
QY 207 AAEAFLSGRGNLSNSRTYRRQNSDNTSRTNWPSVSNPG-TTNTNTSNNSTNSQASQSN 265
Db 273 QNISTINNANNOQATTNMSQPAQPKSSTNADQASQPAHETNSNGTNDKTNESSQSD 332
QY 266 DIDSLLKQLYKLPLSORHVESDGLVFD---PAQITSRARGVAVPHGDHYHFTPYQSMS 321
Db 333 -----VNQYPPADESLQDAIKNPAIIDKEHT-----ADNWRPIDFQMKN 372
QY 322 ELBERIARIIPLYRSNHVWPDSPQSPQPTPEPS-----PGQOPAPNLK----- 368
Db 373 DKGER-----QFYHYA-----STVEPATVIFTKGTPIELGLKTASTWKK 412
QY 369 ---IDSNSLSVLVRKVG---GYV-FEEKGTSRYVFAKDLPSFTVKNLESKLSQESV 421
Db 413 FEVVEGDKLPVELSVSDSKDYAIRFPVSNGTREV-----KIVSSIEYGENIHEDY 465
QY 422 SHTLTAKKENVAPRQDEFYD-KAYN---LITEAHKALFXNKGNSDFQALDKLLERLNDE 477
Db 466 DYTLMWFAQPIITNPPDDYDVBDETYNLQKLLAPVHKA---KTLERQVVELEKLEKLP 521
QY 478 -SINKEKLVDDLAFLA-----PITHEPERLKGNSQIEVTEDEVRIAQLADKYTTSDG 529
Db 522 YKAEYKKKLDQTRVELADQVKSATFENVTPTNDQDLTDQEAHFVFESEENSESVMDG 581
QY 530 YIFDEHDIISD--EGDAYVT-PMHGSHWICKDLSDEKVAQAQYTKERG-----ILPPS 582
Db 582 FV--EHPFYTATLNGQKYVVMKTKDQSYW--KOLIVEGKVTTVTSKPKNSRTLIFFYI 637
QY 583 PDADVKNPTGDSAAAIYNRVKGEKRIPLVRLPMVEHTVEVKNGNLIIPKHDIHNI-- 640
Db 638 PD-----KAVYNAI---VKVVVANIGEGQYHVRINQDINTKDDTSDQNTS 682
QY 641 -----KEAFWD--DHTYKAPNGYTLDELPAITIKYVVEHPDERPHSN---DCMGN 684
Db 683 EPLNVQGTQSGKGVADTDAVENSSTATNPKDASDKADVIE-----PESDVVKDADNN 733
QY 685 ASEHVLGKKDHSDEPNKFNKFADEBPVEETPAEPEVPQVETEKVEBAQLKEAEVLLAKY 744
Db 734 IDKDVQHDVDHLSDMSDNNHFD-----KYDLKEMDTQIAKDTDR 772
QY 745 SLKANATETIAGURNLTL-----QIMDNNSIMAEKLLALLKG-SNPS 788
Db 773 NVDKDADNSV-GMSSNVDTKDSNKNKDKVITQLNHIADKNHHTGKAALKLVVKNYNTD 831
QY 789 SVSKEK 794
Db 832 KVTDKK 837
```

```

/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/961,083
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brookes, A. Anders
/ REGISTRATION NUMBER: 36,373
/ REFERENCE/DOCKET NUMBER: PB340P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 182:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 447 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-961-083-182

Query Match      28.5%; Score 1187.5; DB 3; Length 447;
Best Local Similarity 46.3%; Pred. No. 4.7e-84;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10;

QY 4 LGLYQARTVKNRRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSHG 63
Db 1 LNCQRSQENKNNRVSYVDGSSQSKSENLTDPQVSKQEGIAEQIVIKITDQGYVTSHG 60
QY 64 DHYHYNGKVPYDAIISSELLMKDPNYKLKDEIVNEVGKGYIKVDGKYVYVYLKDAAHA 123
Db 61 DHYHYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVGKGYIKVDGKYVYVYLKDAAHA 120
QY 124 DNVRTKEEINRQKQHSQHREGTTPNDGVALARSQGRYTTDDGYIFNASDIIEDTGA 183
Db 121 DNVRTKDEINRQKQHVKNDE---KVNNSNVAARSQGRYTTNDGYVFNPAIDIEDTGA 176
QY 184 YIVPHGDHGHYIPKQNELSASELAABAFSLGKGNLSNRTYRRQNSDTSRTNWVPSVSN 243
Db 177 YIVPHGCHYHYIPKSDLSASELAABAFSLGKGNLSNRTYRRQNSDTSRTNWVPSVSN 243
QY 244 PGTNTNTNNTNNTNSQAQSQNDIDSLKQLYKLPLSRHVSDGLVFPDPAITSTARG 303
Db 225 ---NTQSAKSTGSKPANKSENQLQSLKELYDSPSAQRYSESDGLVFPDPAKISRTPNG 280
QY 304 VAVPHGDHGHYIPYSQMSLEERARIIPLYRSNHHVDPDSRPEQSPQTPPEPSGPQP 363
Db 281 VAIPLHGDHGHYIPYSKLSALEKIAHWV-----PNEVSSLSLSSNPSS--- 309
QY 364 APNLKIDNSLSLVQLRVKGEYGFVEEKIGRYVFAKDLPSGTVNKLSKLSKQBSVSH 423
Db 310 ----ISGTGSTVSTNAK-----PNEVSSLSLSSNPSS--- 339
QY 424 TLTAKEENVAPRQEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLLERLNDESTNKEK 483
Db 340 -LTTSKELSSASDGYIFNPK-DIVEETATAYIVRHG--DHFHYIPK-----SNQIGQPT 389

```

```

QY 484 LVDDLAFAPITHPERLGPKNQSIQYVTEDEVRIQLADKYTSDGYIFDEHDIISDEGD 543
Db 390 LPNNSLATPSP-SLPINPGTSHKHE-----EDGYGFDPANRIIAEDS 431
QY 544 AYVTPHMGHSHWICK 558
Db 432 GFVMSHGDHNYFFK 446

RESULT 10
US-09-536-784-182
/ Sequence 182, Application US/09536784
/ Patent No. 6573082
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/536,784
/ FILING DATE: 30-Oct-1997
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/961,083
/ FILING DATE: OCT-30-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Michelle S. Marks
/ REGISTRATION NUMBER: 41,971
/ REFERENCE/DOCKET NUMBER: PB340P3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 182:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 447 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match      28.5%; Score 1187.5; DB 4; Length 447;
Best Local Similarity 46.3%; Pred. No. 4.7e-84;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10;

QY 4 LGLYQARTVKNRRVSYIDGKQATKTNLTPEVSKREGINAEQIVIKITDQGYVTSHG 63
Db 1 LNCQRSQENKNNRVSYVDGSSQSKSENLTDPQVSKQEGIAEQIVIKITDQGYVTSHG 60
QY 64 DHYHYNGKVPYDAIISSELLMKDPNYKLKDEIVNEVGKGYIKVDGKYVYVYLKDAAHA 123
Db 61 DHYHYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVGKGYIKVDGKYVYVYLKDAAHA 120
QY 124 DNVRTKEEINRQKQHSQHREGTTPNDGVALARSQGRYTTDDGYIFNASDIIEDTGA 183
Db 121 DNVRTKDEINRQKQHVKNDE---KVNNSNVAARSQGRYTTNDGYVFNPAIDIEDTGA 176
QY 184 YIVPHGDHGHYIPKQNELSASELAABAFSLGKGNLSNRTYRRQNSDTSRTNWVPSVSN 243
Db 177 YIVPHGCHYHYIPKSDLSASELAABAFSLGKGNLSNRTYRRQNSDTSRTNWVPSVSN 243

```



```

;
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-536-784-66

```

Query Match	63.6%;	Score 2649.5;	DB 4;	Length 763;
Best Local Similarity	66.8%;	Pred. No. 2.3e-197;		
Matches 519;	Conservative	77;	Mismatches 132;	Indels 49;
			Gaps 8;	

QY	1	SYELGLYQARTV-KENNRVSYIDGQATQXTENLTDPDEVSKREGINAEQIVIKITDQGV	59
Db	2	SYELGRHQAGQVKESNRVSYIDGQAGQAEMLTPDEVSKREGINAEQXVITIKDQGV	61
QY	60	TSHGDHYHYNGKVPYDALLSEELLMDPNVKLKDEDI NEVKGGVIVKDGKVVYVLKD	119
Db	62	TSHGDHYHYNGKVPYDALLSEELLMDPNYQKDSI VNEIKGGVIVKNGKVVYVLKD	121
QY	120	AAHADNVRTKEEINRQKQHSQHREGCTPRNDGAVALAR SQGRYTTDDCGYIFNADSIIED	179
Db	122	AAHADNRTKEEIKRQKQSRSHNHS--RADNAVAAARQAQGRYTTDDCGYIFNADSIIED	178
QY	180	TGDYAYIVPHGDHYHYIIPKNELASELAAAFISGRGNLSNRTYRQNSDMNTSTNWVP	239
Db	179	TGDYAYIVPHGDHYHYIIPKNELASELAAAFAYNG-----KQGRSPSSSSSYNA	227
QY	240	SVSPNGTTNTNTSNNSNTNSQASQNSDIDSLLLQOLYKPLSOPHVESDGLVFDPAQITSR	299
Db	228	NPAQPRISENHNLVTVTYTHQ-KOGENISSLLRELYAKPLSEKHVESDGLIFDPAQITSR	286
QY	300	TARGVAVPHGDHYHFIPYSQMSLEERIARIIPLRYRSNHWPDSRPEQSPQPTPEPSP	359
Db	287	TARGVAVPHGNHYHFIPYEQMSLEEKRIARIIPLRYRSNHWPDSRPEQSPQSPTEPSP	346
QY	360	GPQAPNULK-IDSN---SSLSVQLVRKGVGYVFEKGISRYVAFKDLPSETVKNLESLK	415
Db	347	SPQAPNPQPAPSNPIDEKVLKVEARVKGVDGYFEENGVSRYVIPA KDL SAETAAGIDSKL	406
QY	416	SKQSVSHITLAKENAVAPDOEFYKAYNLLTEAHKALFXNKGNSDFOALDKILLERIN	475
Db	407	AKQBSLSHLGAKKTYDIPSSDRFPYKAYOLLARIHQDLDNKGROVDFEALDNLESLK	466
QY	476	DESTNKEKLVDDLLAFAPITHPERLGKPNQSIEYTEDEVRIQAQLADKYTTSDGYIFDEH	535
Db	467	DVXSDKVKLVXDILAFAPIRHBERLGKPNQAITYTDDDEIQVAKLAGKYTTEDGYIFDPR	526
QY	536	DIIISDEGDAYTPHMGHSHWIGKDSLSDEKVAQAAYTEKGILLPSPDADVKANPTGDS	595
Db	527	DITSDDEGDAYTPHMTSHWIKKDSLEAEERAAQAAYAKEKGLTTPSTDHQDSGNTAEAK	586
QY	596	AAAYNVRVKGKRIPLVRLPMVTEHTVKNGNLIIPHDKDHYHNIAKFAMFDDHTTKAPNG	655
Db	587	AEAYNVRVKAACKVPLDRMPNLYQTVVEVKNGLSIIIPHVDHYHNIAKFEMFDEGLYEAPKG	646
QY	656	YTLDELFA TI KYVVEHPDERPHSGNDGNGNASHEVLGKXOHSDEPNKNFXADEE-----	709
Db	647	YTLDELLATV KYVVEHPNERPHSDNGFGNASDHWQRNKNQCQADNTQTEKPSSEKQCTEK	706
QY	710	VEET-----PAPFVPVQVETEKVRAQLKEAEVLLAKYTD	743
Db	707	EEETPREEKPOSEKPSKPTPEEPSESESPQVETEKVBEKLEAEDLLGKTOD	763

RESULT 8  
US-09-468-656A-6  
; Sequence 6, Application US/09468656A  
; Patent No. 6582706  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-444  
; CURRENT APPLICATION NUMBER: US/09/468,656A  
; CURRENT FILING DATE: 1999-12-02  
; PRIOR APPLICATION NUMBER: 60/113,048  
; PRIOR FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 484  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-468-656A-6

Query Match	28.9%;	Score 1203.5;	DB 4;	Length 484;
Best Local Similarity	45.8%;	Pred. No. 3e-85;		
Matches	259;	Conservative 77;	Mismatches 120;	Indels 109; Gaps 10
Qy	1	SYELGLTQARTVKENNRSVSYIDGKQATQKTENLTDPDEVSKREGINAEGIVIKITDQGYVT	60	
Db	22	AYALNQRHSEKDNRRVSYVDGSSQSKSENLTIPDQVSQREGTQAEQIVIKITDQGYVT	81	
Qy	61	SHGDHYHYKGVKVPYDAIIISBEELMKDPNYKDKOEDIVNEVKGGYVIKVDGKYVYVLKDA	120	
Db	82	SHGDHYHYKGVKVPYDALFSEELMKDPNYQLKDAIVNEVKGGYIIKVDGKYVYVLKDA	141	
Qy	121	AHADNVTKTEINPKOEHSGHRSQREGGTPRNDGAYALARSGRYITTDGQYIFNASDIIEDT	180	
Db	142	AHADNVTKTEINPKOEHVKNDE-----KVNSNVAVARSGRYITTDGQYVFNPAIIEDT	197	
Qy	181	GDAYIVPHGDHYHYPKNELASASLAAAEAPLSGRGNLSNRTYRRQNSDNTSRTNWWVPS	240	
Db	198	GNAYIVPHGGHYHYPKSDLSASLAAAKAHLAKQWQPSQLSVSSTASDN-----	248	
Qy	241	VSNPGTTNTNTSNNSTNSQASQNSDIDSLIKQLYKPLSORHVESDGLVDPDAQITSR	300	
Db	249	-----NTQSAVGSGTSKPAKNSENLQSLKELYDPSAQRYSBESDGLVDPDAKIIISRT	301	
Qy	301	ARGVAVPHGDHYHYPYSSOMSELEERTARIIPLYRSNHNHWDPSPRPEQSPQPTPEPSPG	360	
Db	302	PNGVAIPHGHDHYHYPYKSLUSALSEKIARWVP-----	333	
Qy	361	POPAPNLKIDNSLSVLSQLRVKVGEGYVFEKGISRYVFAKDLPSETVKNLSEKLSQKES	420	
Db	334	-----ISGTGSTVSTNAK-----PNEVVSSLSGLSSNPSS	363	
Qy	421	VSHTLTAKKENVAPRDQEFYDKAYNLTIAHKALFXNKGNSDRDQALDKLLERLNDESTN	480	
Db	364	-----LTTSKELSSASDGYIFNPK-DIVEETATAVIVRHG-----DHFHYIPK-----	410	
Qy	481	KEKIVDLLLAFLAPITIPERLGLKPNISOIEYTEDEVRIQAADKYTTSQYIFDEHDIISD	540	
Db	411	QPTLPNNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFADNRILAE	452	
Qy	541	EGDAYVTPHMGHGMWIGKDSLSDKE	565	
Db	453	DESGFVMSHGHDHNYHFFKKDLITEQ	477	

RESULT 9  
US-08-961-083-182  
; Sequence 182, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:

	QY	596	AAAIYNRVKGEKRIPLVRLPYMVEHTVEXXGNLIIPHKHGYHNLIKFAWFDHDTTKAPNG	655
	Db	607	AEAIIYNRVKAACKVPIDRMPYNLYTVBVKNGSLIIPHYDHYHNLIKFEWFDEGLYEAPKG	666
	QY	656	YTLEDLFATIKYYUHHDPDRPHSDNCWGNASEHVLGK-----KDHS-----	697
	Db	667	YTLEDLLATVKYYYEHHPNERSHSNDGFGNASHVRKNKVDDQSKDFEDKEHDDEVSEPETHP	726
	QY	698	-----DNKNKPADPEEPPVEETPAEPVQPQVETEVAQLKEAEVL	738
	Db	727	ESDEKENHAGLNPSADNLKYESTDTTEETEEBAEDTTDEAIEPOVENSVINAKIADEALL	786
	QY	739	AKVTDSLSKANATETFLAGLRNNLTLOIMNNSIMAAEKLLALLKGSNPSSV	790
	Db	787	EKVTDPSIRQNAMETLTLGKKSLLGTCKDNNTISAEVDSSLALLKESOPAPI	838
 RESULT 6 US-08-961-083-66 ; Sequence 66, Application US/08961083 ; Patent No. 6159469 ; GENERAL INFORMATION: ; APPLICANT: Choi et. al. ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines ; NUMBER OF SEQUENCES: 452 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: Human Genome Sciences, Inc. ; STREET: 9410 Key West Avenue ; CITY: Rockville ; STATE: Maryland ; COUNTRY: USA ; ZIP: 20850 ; COMPUTER READABLE FORM: ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage ; COMPUTER: HP Vectra 486/33 ; OPERATING SYSTEM: MSDOS version 6.2 ; SOFTWARE: ASCII Text ; CURRENT APPLICATION DATA: ; FILING DATE: ; PRIORITY APPLICATION NO.: US/08/961,083 ; CLASSIFICATION: 435 ; PRIOR APPLICATION DATA: ; APPLICATION NUMBER: ; FILING DATE: ; ATTORNEY/AGENT INFORMATION: ; NAME: Brookes, A. Anders ; REGISTRATION NUMBER: 36,373 ; REFERENCE/DOCKET NUMBER: PB340P2 ; TELECOMMUNICATION INFORMATION: ; TELEPHONE: (301) 309-8504 ; TELEFAX: (301) 309-8512 ; INFORMATION FOR SEQ ID NO: 66: ; SEQUENCE CHARACTERISTICS: ; LENGTH: 763 amino acids ; TYPE: amino acid ; STRANDEDNESS: single ; TOPOLOGY: linear ; MOLECULE TYPE: protein ; JS-08-961-083-66				

Qy	120	A A H A D N V R T K E E I N R Q O E H S Q H R E G G T P R N D G A V A L A R S O G R Y T T D D G Y I F N A S D I I L D	179
Db	122	A A H A D N I R T K E E I K R Q O E S H N H N S - - - R A D N A V A A A Q C R Y T T D D G Y I F N A S D I I L D	178
Qy	180	T G D A Y I V P H G D H V H Y I P K N E L S A S E L A A A F L S G R G N L S N S R Y R R O N S D N T S R T N M V P	239
Db	179	T G D A Y I V P H G D H V H Y I P K N E L S A S E L A A A A Y W N G - - - - - K O G S R P S S S S Y N A	227
Qy	240	S V S N P G T T N T S N N S N T N S Q A G S N D I D S L L K O L Y K L P L S Q R H V E S D G L V F D P A Q I T S R	299
Db	228	N P A Q P R L S E N H N I T V T T Y H Q - N O G E N I S L L R E L Y A K P L S E R H V E S D G L I F D P A Q I T S R	286
Qy	300	T A R G V A V P H G D H V H F I P Y S O M S E L B E R I A R I I P U R Y S N H W V P D S R P O S P O P P E S P	359
Db	287	T A R G V A V P H G N H V H F I P Y O M S E L E K I A R I I P U R Y S N H W V P O S R P P O S P O S T P E S P	346
Qy	360	G P O P A P N L K - I D S N - - - S S I V S Q L V R V G E G Y V F E E K G I S Y V F A K O J L P S T V T N K L S K L	415
Db	347	S P Q P A P N P Q A P S N P I D E K L V K E A V R K V G D G Y V F E E N G V S R Y I P A K D L S A E T A A G I D S K L	406
Qy	416	S K O E S V H T L T A K K E N V A P R D O F Y D K A Y N L L T A H K A L F X N K G N S D F O A L D K L L E R I N	475
Db	407	A K O E S L S H K L G A K K T D L P S S D R E F Y N K A Y D L L A R I H Q D L L D N K G Q V D F A L D N L L E R I K	466
Qy	476	D E S T N E K L V D D L L A F L A P I T H P R L G K P N S O I E Y T E D E V R A I O A L A D K Y T T S D G Y I F D E H	535
Db	467	D V S Z K V K L V X D I L A F L A P I R H P R L G K P N A Q I T T D D E I O V A K L A G K Y T T D G Y I F D P R	526
Qy	536	D I I S D E G D A Y V T P H M G S H W I G K D S D K E K A V A A Q A Y T K E G I L P P S P D A D V K A N P T G D S	595
Db	527	D I T S D E G D A V V T P H M T H S H W I K K D S L S E A R A A A Q A Y A K E K G L T P P S T D H Q D S G N T E A K G	586
Qy	596	A A A I N Y N K G E K R I P L V R L P M V E H T V E V K N G N L I I P H K D H V H N I K F A W F D D H T Y K A P N G	655
Db	587	A E A I N Y N K A A K V P L D R M P Y N L Q Y T V E V K N G S L I I P H Y D H Y H N I K F E W F D E G L Y E A P K G	646
Qy	656	Y T L E D L P A T I K Y Y V E H P D R P H S N D G N G N A S E H V L G K K D H S D P N K N F K A D E E - - - - - P	709
Db	647	Y T L E D L L A T V K Y Y V E H P E N R P H S D N G F G N A S D H V Q R N K N G Q A D T N Q T E K P S E E K P Q T E K P	706
Qy	710	V E E T - - - - - P A P E V P Q V E T E K Y E A Q L K E A E V L L A K V T D	743
Db	707	E E E T P R E E K P Q S E K P S P K P T E E P E S P E S E E P Q V E T E K Y E E K L R E A E D L L G K I Q D	763
RESULT 7			
US-09-536-784--66			
; Sequence 66, Application US/09536784			
; Patent No. 6573082			
; GENERAL INFORMATION:			
; APPLICANT: Choi et. al.			
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines			
; NUMBER OF SEQUENCES: 452			
; CORRESPONDENCE ADDRESS:			
ADDRESS: Human Genome Sciences, Inc.			
STREET: 9410 Key West Avenue			
CITY: Rockville			
STATE: Maryland			
COUNTRY: USA			
ZIP: 20850			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage			
COMPUTER: HP Vectra 486/33			
OPERATING SYSTEM: MSDOS version 6.2			
SOFTWARE: ASCII Text			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/536,784			
FILING DATE: 30-Oct-1997			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 08/961,083			
FILING DATE: OCT-30-1997			
ATTORNEY/AGENT INFORMATION:			



```

QY 1 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
DB 1 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYYVYLKDA 120
DB 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYYVYLKDA 120
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180
DB 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180
QY 181 GDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 181 GDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
QY 241 VSNPGTTNTNTSNNSNTNSQASQNDIDSLLKQYKPLSQRHVESDGLVDFDPAQITST 300
DB 241 VSNPGTTNTNTSNNSNTNSQASQNDIDSLLKQYKPLSQRHVESDGLVDFDPAQITST 300
QY 301 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPSRPPQSPQPTPEPSPG 360
DB 301 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPSRPPQSPQPTPEPSPG 360
QY 361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQES 420
DB 361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQES 420
QY 421 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFYXNKGNSDFQALDKLLERLNDESTN 480
DB 421 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFYXNKGNSDFQALDKLLERLNDESTN 480
QY 481 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVETRIQALADKYTTSDGYIFDEHDIISD 540
DB 481 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVETRIQALADKYTTSDGYIFDEHDIISD 540
QY 541 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 600
DB 541 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 600
QY 601 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNIKFADFDDHTTYKAPNGYTTLED 660
DB 601 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNIKFADFDDHTTYKAPNGYTTLED 660
QY 661 LFATIKYVVEHPDERPHSDNGWGNASHVILGKXDSDPNKNFKADEEPVEETPAEPEVP 720
DB 661 LFATIKYVVEHPDERPHSDNGWGNASHVILGKXDSDPNKNFKADEEPVEETPAEPEVP 720
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
DB 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
QY 781 LLKGSNPSSVSKKIN 796
DB 781 LLKGSNPSSVSKKIN 796

```

```

RESULT 3
US-09-468-656A-8
; Sequence 8, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; PRIOR APPLICATION NUMBER: 1999-12-02
; PRIOR FILING DATE: 1998-12-21

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-8

Query Match 100.0%; Score 4163; DB 4; Length 819;
Best Local Similarity 99.9%; Pred. No. 5.1e-315; Indels 0; Gaps 0;
Matches 795; Conservative 0; Mismatches 1;

QY 1 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
DB 21 SYELGLYQARTKNNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYYVYLKDA 120
DB 81 SHGDHYHYNGKVPYDAIISEELLMDPNYKLDKDEDIVNEVKGYYIKVDGKYYVYLKDA 140
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 180
DB 141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGVALARSQRYTTDDGYIFNASDIIEDT 200
QY 181 GDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 240
DB 201 GDAYIVPHGDHYHYIPKNELASASELAFLSGRGNLSNRTYRRQNSDNTSRTNWVPS 260
QY 241 VSNPGTTNTNTSNNSNTNSQASQNDIDSLLKQYKPLSQRHVESDGLVDFDPAQITST 300
DB 261 VSNPGTTNTNTSNNSNTNSQASQNDIDSLLKQYKPLSQRHVESDGLVDFDPAQITST 320
QY 301 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPSRPPQSPQPTPEPSPG 360
DB 321 ARGVAVPHGDHYHYIPYSQMSSELEERTARIIPLYRSNHWVPSRPPQSPQPTPEPSPG 380
QY 361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQES 420
DB 381 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQES 440
QY 421 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFYXNKGNSDFQALDKLLERLNDESTN 480
DB 441 VSHTLTAKKENVAPRQDFYDKAYNLLTEAHKALFYXNKGNSDFQALDKLLERLNDESTN 500
QY 481 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVETRIQALADKYTTSDGYIFDEHDIISD 540
DB 501 KEKLVDDLLAFLAPITHPERLGPNSQIETEDVETRIQALADKYTTSDGYIFDEHDIISD 560
QY 541 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 600
DB 561 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEGILPPSPDADVKANPTGDSAAIY 620
QY 601 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNIKFADFDDHTTYKAPNGYTTLED 660
DB 621 NRKVGKRIPLVRLPYMVEHTEVKGNNLIIPKHQHYHNIKFADFDDHTTYKAPNGYTTLED 680
QY 661 LFATIKYVVEHPDERPHSDNGWGNASHVILGKXDSDPNKNFKADEEPVEETPAEPEVP 720
DB 681 LFATIKYVVEHPDERPHSDNGWGNASHVILGKXDSDPNKNFKADEEPVEETPAEPEVP 740
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
DB 741 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 800
QY 781 LLKGSNPSSVSKKIN 796
DB 801 LLKGSNPSSVSKKIN 816

```

```

RESULT 4
US-09-468-656A-10
; Sequence 10, Application US/09468656A
; Patent No. 6582706

```

```
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-56

Query Match 100.0%; Score 4163; DB 3; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.9e-315;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENRYSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGVVT 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1 SYELGLYQARTVKENRYSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGVVT 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDI VNEVKGGYVIKVDGKYVYLKDA 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLKDEDI VNEVKGGYVIKVDGKYVYLKDA 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 AHADNVRTKEEINRQKSHQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 AHADNVRTKEEINRQKSHQHREGGTPRNDGAVALARSGRYTTDDGYIFNASDIIEDT 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 GDAYIVPHGDHYHYIPKNELSASELAAAEFLSGRGNLSNRYRQNSDNTSRINWVPS 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 GDAYIVPHGDHYHYIPKNELSASELAAAEFLSGRGNLSNRYRQNSDNTSRINWVPS 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 VSNPGTTNTSNNTNSQASQSDNDISLLKQLYKPLSRHVSDGLVDPQAITSRT 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 VSNPGTTNTSNNTNSQASQSDNDISLLKQLYKPLSRHVSDGLVDPQAITSRT 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 ARGVAVPHGDHYHYIPYSCMSELEERARIIPLRYSNHNWVDSRPEQSPQPTPEPSG 360
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 ARGVAVPHGDHYHYIPYSCMSELEERARIIPLRYSNHNWVDSRPEQSPQPTPEPSG 360
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 PQAPNLIKIDNSSLSVQLVRKVGYYVFEKIGSYVFAKOLPSETVKNLESKLSKQES 420
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 PQAPNLIKIDNSSLSVQLVRKVGYYVFEKIGSYVFAKOLPSETVKNLESKLSKQES 420
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 VSHTLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 481 KEKLVDDLLAFADITHPERLGKPN SQIEYTEDEVRIAQLADKYTTSQGYIFDEHDIISD 540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKGILPPSPDADVKANPTGDSAAAIY 600
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKEKGILPPSPDADVKANPTGDSAAAIY 600
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 601 NRKGEKRIPLVRLPYMVVEHTVEVKNGLIIPHKDHYHNIKFAWFDHDTYKAPNGYTTLED 660
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 661 LFAIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNFKADBEPPVETPAEPV 720
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 781 LLKGNPSSVSSEKIN 796
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; US-09-536-784-56
; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
; US-09-536-784-56

Query Match 100.0%; Score 4163; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 4.9e-315;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



Qy 652 APNGYTLEDFATIKYVVEHPDE 674  
Db 777 -----EHEDE 781

Search completed: October 1, 2004, 07:16:43  
Job time : 92 secs





Db 330 KDI-----DLSKEIPNPH-----HDDHEDHKGHHHDSBHK----- 360  
QY 461 NSDFQALDKLLERLNDSTNKEKLVDD---LLAFLAPITHPERLGKPNQSIQIYETDEVDRI 517  
Db 361 -----HEEHEHTKSNKLSDEQKGLIYLA-----EKLGLNPNQIEV----- 396  
QY 518 AQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGHSHW-----IGK-----DLSLD 563  
Db 397 -----LTSEDSIIFKY-----PHDHSHTIASKDIKPIPDGHHDSHA 438  
QY 564 KEKVAQAQYTKKKG-----IL---PPSPDADVKANP---TGDSAAAAYNRVKGKRIIP 610  
Db 439 KDKV-GMATLKQIGFDDBIQDILHADAPTTPPSNETNPEKMRQWLATVTKINIGQRTNP 497  
QY 611 LVRLPYMVEHTVVEKGNLIIIPHKD-----HYHNKIFAW-----FDDHTY--KAP----- 653  
Db 498 FQRFGLSLMPNIEVL-GIGFTPIINDMTVPLOFKLQKLMWNTGITDYSFLDKMPLLEGL 556  
QY 654 ---NGYLEDLFATIKY-----YVEHP-----DERPHS 678  
Db 557 DISQNG--IKDLSFLTKYKQLSLTAAANGITSLKPLAELNQLVLVSHNNISDLTPLS 614  
QY 679 N-----DGMGNASEHVLGKHSE-DPNKNFKADEEPVEETPAE----- 716  
Db 615 NLTKLQELHLDHNNVKNLSA-LSGKKDLKVLDSNNKSADLSTLTKTTSLETLNLTNTS 673  
QY 717 -----PEVPQV-----ETEKVEAQLKEAEVLLAKVTDSSLK----- 747  
Db 674 NLSFLKQNPVSNITINNNAKLASLDGIEBSEDIKVEAEAGNQIKSLVLNKGSLKFLNV 733  
QY 748 -ANATETIAGLRNNLTQIOMDNNSIMAEAEKLLALLKGNSSPSVSKEKI 795  
Db 734 TNNQILSLEGVNNYTSLETL-----SVSKNKL 760

RESULT 23  
Q8E029  
ID Q8E029 PRELIMINARY; PRT; 877 AA.  
AC Q8E029;  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Protein of unknown function/lipoprotein, putative.  
GN SAG0907.  
OS Streptococcus agalactiae (serotype V).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=216466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2603 V/R / Serotype V;  
RX MEDLINE=22222988; PubMed=12200547;  
RA Tettelein H., Masiugnani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,  
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative genomic analysis of an  
RT emerging human pathogen, serotype V Streptococcus agalactiae";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
DR EMBL; AE014233; AAM99793.1; -;  
DR TIGR; SAG0907; -;  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR007092; LRR SDS22.  
DR InterPro; IPR006270; Strp\_his\_triad.  
DR Pfam; PF00560; LRR; 3.  
DR Pfam; PF04270; strep\_his\_triad; 4.  
KW Lipoprotein; Complete proteome.  
SQ SEQUENCE 877 AA; 97563 MW; 4829E9E593A4F27 CRC64;

Query Match 7.3%; Score 302; DB 16; Length 877;  
Best Local Similarity 20.8%; Pred. No. 1.4e-08;  
Matches 200; Conservative 95; Mismatches 263; Indels 402; Gaps 43;  
QY 13 KENNRVYIDGKQATOKTENLTDPDEV---SKREGI-----NAEQIVIKITD 55  
Db 26 QEHKNSHHIKTKQVAKKANKKKVSVKESHKKRGVAGVDPTDDGELLTKDKSKILSHD 85  
QY 56 QGYVTSGDHYHYNGKVPYDAIISEELLMKDPNKLKDEIVNEVKGYYIKVDGKYYV 115  
Db 86 SGIVVAHGNHSHFT-----FYSDLKSGKFS 110  
QY 116 YLKDAAHADNVRTKEEINRQKQHSQHREGTTPNDGAVALARSQRYTTDDGVIENASD 175  
Db 111 YL-----IPNANTKTNKQAVRNPKA-----GAVAV-----NTLNDGVVFNPAD 149  
QY 176 IIEDTDAYIVPHGDHYHYIPKNELSASELAFAFLSGRNLNSRTYRQNSDNTSRT 235  
Db 150 IVSEDAHYVVRGDFHYIPKASLSQKQVQASEAVSRGLGNQNS--HYRVNSKSIAGL 207  
QY 236 NWPSVSNPGTTNTNTNNSNTNSQASQNDIDSLKQLYKPLSQRHVESDGLVFPDPAQ 295  
Db 208 HY-PT-----SNGFLFNGRG 221  
QY 296 ITSETARGVAVPHGDHYHYTPYSOMSELEBRIARIIPLRVRSNHM-----VPDS 344  
Db 222 IKGTTPTGILVEHNNHILHIFISFADL-----KKGWGSIAIDRYQPKKADS 266  
QY 345 RPEQSPQPTPEPSPGPPQPNLKNIDNSSIVSOLVRKV-----GE-GYVF 389  
Db 267 KKQSPS-----SKPRTENTLPKIDKDL-AVLARELHLDISIRIVLKTNGEIGFEY 318  
QY 390 EEKIGSYRVFANOLPSETVKNLESKLSQBSVSHITLAKKENVAPROQEFVDKAYNLITE 449  
Db 319 PHDDHTHVIWAKDI-----DLSKPIPNPH-----HDDHEDHKGHHHDS 357  
QY 450 AHKALFXNKGKNSDFQALDKLLERLNDSTNKEKLVDD---LLAFLAPITHPERLGKPN 506  
Db 358 DHK-----HEEHTKSNKLSDEQKGLIYLA-----EKLGLNPN 392  
QY 507 QIYETDEVRIAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGHSHW-----IGK- 558  
Db 393 QIEV-----LTSEDSIIFKY-----PHDHSHTIASKDIKPI 427  
QY 559 -----DLSDKKEKVAQAQYTKKKG-----IL---PPSPDADVKANP---TGDSAAA 599  
Db 428 IPDGHHDHSHAKDKV-GMATLKQIGFDDBIQDILHADAPTTPPSNETNPEKMRQWLATV 486  
QY 600 YNRVKGKRIPLVRLPYMVEHTVVEKGNLIIIPHKD-----HYHNKIFAW-----FDDHT 649  
Db 487 TKINIGQRTNPFQFGLSLMPNIEVL-GIGFTPIINDMTVPLOFKLQKLMWNTGITDYS 545  
QY 650 Y--KAP-----NGYLEDLFATIKY-----YVEHP----- 672  
Db 546 FLDKMPLEGLDISQNG--IKDLSFLTKYKQLSLTAAANGITSLKPLAELNQLVLVLS 603  
QY 673 -----DERPHS-----DGMGNASEHVLGKHSE-DPNKNFKADEEPVEETPA 715  
Db 604 HNNISDLTPLSLNLTQLOELYLDHNNVKNLSA-LSGKKDLKVLDSNNKSADLSTLTKTTS 662  
QY 716 E-----PEVPQV-----ETEKVEAQLKEAEVLLAK 740  
Db 663 ETLNLTNTNLSFLKQNPVSNITINNNAKLASLDGIEBSEDIKVEAEAGNQIKSLVLK 722  
QY 741 VTDSSLK-----ANATETIAGLRNNLTQIOMDNNSIMAEAEKLLALLKGNSSPSVSKEKI 795  
Db 723 NKQGLKFLNVNNTNQLTSLEGVNNYTSLETL-----SVSKNKL 760

RESULT 24  
Q8DQ06  
ID Q8DQ06 PRELIMINARY; PRT; 182 AA.





```

Db 81 YVTHSGDHVHYFNGKVPYDAIISEELMTDPNHYFKQSDVINELDGYVIVKNGNYYVYL 140
QY 118 KDAHADNVRTKEEINRQKQEHSHQ-REGG-----TPRNDGAVALARSGRYTTDDGY 169
Db 141 KPGSKRNIRTKQIAEQAQVAKGKEAKGLAQVAHLSKEEVAANEAKRGQRYTTDDGY 200
QY 170 IFNASDIIEDGTDAYIVPHGDHVIYIPKNELSASELAARAFSL--GRGNLSNRTYR 226
Db 201 IFSPTDIIIDLDGAYLVPHGNHVIYIPKDLSELAARAFSL--GRGNLSNRTYR 258
QY 227 QNSDNTSRNWNVPVS-NEFTNTNTSN-----NSNTNSQASQND-----IDSLK 272
Db 259 TPAPGRKAP-IPDVTNPGQHPDNGGYPHAPPNDASQNKHQDEFKGTFFKELL 317
QY 273 QLYKLPQSORHVESGLVDPQAQITRTARGVAVPHGDHVIYIPYQMSLEERARIIP 332
Db 318 QLRHLDKRVHVEEDGLIPEPTQVKSNAFYVPHGDHVIYIPRSQLSPELMELAD--- 374
QY 333 LRYRSHNHWVDSRPEQSPQTPPEPSPGPAPNKLKIDSNLSVLSQVRKVGEGYVFEK 392
Db 375 -RYLAG-----QTDNDN-----GSDH----- 390
QY 393 GISRYVFAKDLPSFTVKNLESKQESVSHLTAKKENVAPRQDQFYDKAYNLLTEAHK 452
Db 391 -----SKPSDKE-VTHTFGLHR-----IKAY----- 410
QY 453 ALFXNKGNSDFQALDKLLERLNDESTNKEKLVDDLAFAPITHPERLGKPNQSEIYTE 512
Db 411 -----GKGLD-----GKP----- 418
QY 513 DEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAQ 572
Db 419 -----YDTSAYVFSKESIHSVDKSGVTAKGDFHVIYIGFGELEQYELDEVANW 467
QY 573 TKEKG-----ILPPSPDADVKANPTGDSAAA-----IYNR----- 602
Db 468 VKAQGADELVAALDQEQGKEKPLFTTKVSRKVTGDKGVYIMPKDQGYFYARYQLDL 527
QY 603 -----VKGEK-----IPLVRLPYMVEHTVEVKNGLIIPKH 636
Db 528 TQIAFAEQLMLKDKKHRYRDIIVDTGIEPLAVDLSSPLMHAGNATYDTGSSFVIDH 587
QY 637 YHNIKFAWFDHTYKAPNGYTTLEDLPATIKYVVEHDPHPSNDG-----GNASEH 688
Db 588 IHVVPYSWL-----TFNQIATIKYVMOHPEVRP---DWSKPGHESSGVSIPN 632
QY 689 VLKGDHSDPNKVFKADEPVEETPAE-----PE-----VP 720
Db 633 VTPLDKRAGPNMQIIHSAEVOKALAEGRFAAPDGYIFDPRDLAKETFWVKDGSFIP 692
QY 721 QVE-----TEKVEAQLKEAEVLAKVTDSSKANATETLAGLNNLTQIMDNNSI 771
Db 693 RADGSSLRINKSDLSQAEWQQQELLAKN-----AGDATDT-----DKPEE 735
QY 772 MAEAEKLLALLKGNPSFVSSEK 794
Db 736 KQQAQK---SNENQQPSEASKER 755

```

RESULT 18

```

Q9ZHG7
ID Q9ZHG7 PRELIMINARY; PRT; 822 AA.
AC Q9ZHG7;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=R268;
RX MEDLINE=99115568; PubMed=9916102;
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA Schmitz N., Iuetzchen R., Podbielski A.;
RT "Lmb, a protein with similarities to the Irai adhesin family, mediates
RL attachment of Streptococcus agalactiae to human laminin.";
RL Infect. Immun. 67:871-878 (1999).
DR EMBL: AF062533; AAD13797.1; --
DR PIR: T46758; T46758; Strep_his triad.
DR InterPro: IPR006270; Strep_his triad.
DR Pfam: PF04270; Strep_his triad; 6.
DR TIGRFAMs: TIGR01363; Strep_his triad; 4.
KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 22.4%; Score 934; DB 2; Length 822;
Best Local Similarity 29.4%; Pred. No. 5.3e-44;
Matches 271; Conservative 99; Mismatches 318; Gaps 32;

QY 1 SYELGLYQARTVKENNRVSVID--GKQATQKTNLTDPDEVSKREGINAEQIVIKITDQG 57
Db 22 SYVLGKHMLATKQIAIYIDDSKGVKAPKT-NKTMDOISAEGLISAEQIVVYKITDQG 80
QY 58 YVTHSGDHVHYFNGKVPYDAIISEELMKDPNYKLKDEIVNEVKGYVIVKDGKYYVYL 117
Db 81 YVTHSGDHVHYFNGKVPYDAIISEELMTDPNHYFKQSDVINELDGYVIVKNGNYYVYL 140
QY 118 KDAHADNVRTKEEINRQKQEHSHQ-REGG-----TPRNDGAVALARSGRYTTDDGY 169
Db 141 KPGSKRNIRTKQIAEQAQVAKGKEAKGLAQVAHLSKEEVAANEAKRGQRYTTDDGY 200
QY 170 IFNASDIIEDGTDAYIVPHGDHVIYIPKNELSASELAARAFSL--GRGNLSNRTYR 226
Db 201 IFSPTDIIIDLDGAYLVPHGNHVIYIPKDLSELAARAFSL--GRGNLSNRTYR 258
QY 227 QNSDNTSRNWNVPVS-NEFTNTNTSN-----NSNTNSQASQND-----IDSLK 272
Db 259 TPAPGRKAP-IPDVTNPGQHPDNGGYPHAPPNDASQNKHQDEFKGTFFKELL 317
QY 273 QLYKLPQSORHVESGLVDPQAQITRTARGVAVPHGDHVIYIPYQMSLEERARIIP 332
Db 318 QLRHLDKRVHVEEDGLIPEPTQVKSNAFYVPHGDHVIYIPRSQLSPELMELAD--- 374
QY 333 LRYRSHNHWVDSRPEQSPQTPPEPSPGPAPNKLKIDSNLSVLSQVRKVGEGYVFEK 392
Db 375 -RYLAG-----QTDNDN-----GSDH----- 390
QY 393 GISRYVFAKDLPSFTVKNLESKQESVSHLTAKKENVAPRQDQFYDKAYNLLTEAHK 452
Db 391 -----SKPSDKE-VTHTFGLHR-----IKAY----- 410
QY 453 ALFXNKGNSDFQALDKLLERLNDESTNKEKLVDDLAFAPITHPERLGKPNQSEIYTE 512
Db 411 -----GKGLD-----GKP----- 418
QY 513 DEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSDEKVAQAQ 572
Db 419 -----YDTSAYVFSKESIHSVDKSGVTAKGDFHVIYIGFGELEQYELDEVANW 467
QY 573 TKEKG-----ILPPSPDADVKANPTGDSAAA-----IYNR----- 602
Db 468 VKAQGADELVAALDQEQGKEKPLFTTKVSRKVTGDKGVYIMPKDQGYFYARYQLDL 527
QY 603 -----VKGEK-----IPLVRLPYMVEHTVEVKNGLIIPKH 636
Db 528 TQIAFAEQLMLKDKKHRYRDIIVDTGIEPLAVDLSSPLMHAGNATYDTGSSFVIDH 587
QY 637 YHNIKFAWFDHTYKAPNGYTTLEDLPATIKYVVEHDPHPSNDG-----GNASEH 688
Db 588 IHVVPYSWL-----TFNQIATIKYVMOHPEVRP---DWSKPGHESSGVSIPN 632
QY 689 VLKGDHSDPNKVFKADEPVEETPAE-----PE-----VP 720

```

RESULT 16					
Q8K5Q1	PRELIMINARY;	PRT;	823 AA.		
ID	Q8K5Q1				
AC	Q8K5Q1;				
DT	01-OCT-2002 (TrEMBLrel. 22, Created)				
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)				
DE	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)				
DE	Histidine triad protein.				
GN	SPYM3_1724.				
OS	Streptococcus pyogenes (serotype M3).				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=198466;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MGAS315 / Serotype M3;				
RX	MEDLINE=2213308; PubMed=12122206;				
RA	Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,				
RA	Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,				
RA	Campbell D.S., Smith I.M., McCormick J.K., Leung D.Y.M.,				
RA	Schlievert P.M., Musser J.M.;				
RT	"Genome sequence of a serotype M3 strain of group A Streptococcus:				
RT	phage-encoded toxins, the high-virulence phenotype, and clone				
RT	emergence.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).				
DR	EMBL; AE014169; AM80331.1;-;				
DR	InterPro; IPR006270; Strep_his_triad.				
DR	Pfam; PF04270; strep_his_triad; 6.				
DR	TIGRFAMs; TIGR01363; strep_his_triad; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 823 AA; 92431 MW; 8C6CBD517A2DD616 CRC64;				
 Query Match 22.6%; Score 940; DB 16; Length 823; Best Local Similarity 30.0%; Pred. No. 2.4e-44; Matches 262; Conservative 122; Mismatches 272; Indels 216; Gaps 30.					
Qy	1 SYELGLVQARTVKENRVSYD---GKQATOKTENLPDEVSKREGINAEQIVIKITDQG 57				
Dd	22 SYQLGKHMHGLATKDNOIAYIDDSKGAKAPKT-NKTMDOQISAEEGISAQIVVKITDQG 80				
Qy	58 YVTSHGDHYHYNGKVPDYDAIIISELLMKDPNYKLKDEDIVNEKVGGYVIKVDGKYVYL 117				
Dd	81 YVTSHGDHYHYNGKVPDYDAIIISELLMTDPNYHFKGSDVINELTDGVIVKNGNYVI 140				
Qy	118 KDAAHADNVRTKEBINQKQEHSH-REG-----TPRNDGAVALARSQRYTTDGY 169				
Dd	141 KPQSCKRNIRTKQIQIAQVAKGTKEAKEGLAQVAHLSKEEVAANEAKRGQRYTTDGY 200				
Qy	170 IFNASDIIEDTGDYIVPHGDHYHIIPKNELSASELAAEAFLS---CRGNLSNRTYRR 226				
Dd	201 IFSPDIIDDLGDYLVPHGNHYHIPKKOLSPSELAQAAYWSQKQRG--ARPSDYRP 258				
Qy	227 QNSDNTSRTNWPVS-VPGTNTNTSN-----NSNTNSQAQSQND-----IDSLLK 272				
Dd	259 TPAPGRKAP-IPDVTPNGCGHQPDNGGYHPAPRPNDASQNKHORDEPKGTFKELL 317				
Qy	273 QLKKPLPUSQRHVESDGLVFDAQITSRTARGVAVPHGDHYHIFPYSQMSELEERARIIP 332				
Dd	318 QLHRDLDKYRHVEDGLIFEPTQVIKSNFAGVVYPHGHDHYHIIPRSQLSPELEMAD--- 374				
Qy	333 LYRESNHWPDSREQSPQPSTPFSPCPQAPNLKIDSNSLSVQLVKVGEGYVFEK 392				
Dd	375 -RYLAGQ-----TEDNDSGSDHSKSDKEVTHFTFLGHRISKAYGKGLDGKPY 419				
Qy	393 GIS-RYYFAKDLPSSETVNKLESKSQESVSHTLTAKKENVAPRDQEFYDKAYMLTEAH 451				
Dd	420 DTSDAYVFSK----ESIHSV-----KSGVTAKGCDHFHYIFGELEQYE 460				
Qy	452 KAUFYN----KGRNSDFQALDKLLERNDESINKELVDDLAFLLAPIITHPERLGKPNQ 507				
Dd	461 LDEVANVWKAKG-----ADELAALADDOEGKPKLFD----- 493				

Qy	508	I EYTED E V R I A Q L A D K Y T T S G Y I F D E H D I I S D E G D A V V T P H M G H S H W I G K D S L S D K E V	56			
Db	494	-----T K Y S R K V T K D G V-----G Y I M P K D G K D Y F Y A R D Q L D-----L	527			
Qy	568	A A Q A Y T K E G I L P P S P D A D V K A N P T G D S A A A I Y N R V K G E K R I P L V R L P V M V E H T V E V K N G	627			
Db	528	T Q I A F A Q E L M L K D K N H Y R I D V T G-----I E P R L A V D V S S L P M H A G N A T Y D T G S	578			
Qy	628	N L I I P H K D H Y H N I K P A W F D D H T Y K A P N G Y T L E D L F A T I K Y Y V E H P D E R P H S N D G M-----	682			
Db	579	S F V I P H I D H I H V P Y S M L-----T R D Q I A T I K Y V M Q H P E V R P-----D V M S K P G H	623			
Qy	683	---G N A S E H V L G K D H S E D P K N K A D E E P V E T P A E-----P E-----	718			
Db	624	E E S G S V I P N V T P L D K R A G M P N W Q I I H S A B E Y Q K A L A E G R F A T P D G Y I F D P R D V L A K E T F V	683			
Qy	719	-----V P O V E-----T E K V E A Q L K E A E V L L A K V T D S S L K A N A T E T L A G L R N L I T	762			
Db	684	W K D G S F I P R A D G S S L R T I N K S D I S Q A E W Q Q A Q E L L A K N-----A G D A T D T-----	730			
Qy	763	L Q I M D N N S I M A E K L I A L L K G S N P S S V S K E K	794			
Db	731	---D K P E K Q Q A D K---S N E N Q Q P S E A S K E E	755			
RESULT 17						
Q8DZ81						
ID	Q8DZ81	PRELIMINARY;	PRT; 822 AA.			
AC	Q8DZ81;					
DT	01-MAR-2003 (TremBLrel. 23, Created)					
DT	01-MAR-2003 (TremBLrel. 23, Last sequence update)					
DT	01-OCT-2003 (TremBLrel. 25, Last annotation update)					
DE	Streptococcus histidine triad family protein.					
GN	SAG1233.					
OS	Streptococcus agalactiae (serotype V).					
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;					
OC	Streptococcus.					
NCBI_TaxID=216466;						
LN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=2603 V/R / Serotype V;					
RX	MEDLINE=2222988; PubMed=12200547;					
RA	Tettelin H., Maignan V., Cieslewicz M.J., Eisen J.A., Peterson S.,					
RA	Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,					
RA	Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,					
RA	DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,					
RA	Radune D., Fedorova N.B., Scanlan R., Khouri H., Mulligan S.,					
RA	Carty H.A., Clive R.T., Van Aken S.E., Gill J.J., Scarbelli M., Mora M.,					
RA	Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,					
RA	Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,					
RA	Fraser C.M.;					
RT	"Complete genome sequence and comparative genomic analysis of an					
RT	emerging human pathogen, serotype V Streptococcus agalactiae.";					
RL	Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).					
RL	EMBL; AE014248; AAN00111.1; .					
DR	TIGR; SAG1233; .					
DR	InterPro; IPR006270; Strep_his_triad.					
DR	Pfam; PF04270; strep_his_triad; 6.					
DR	TIGRFAMS; TIGR01363; strep_his_triad; 1.					
KW	Complete proteome.					
SQ	SEQUENCE 822 AA; 92400 MW; 8CC88DF316727F98 CRC64;					
Query Match 22.5%; Score 937; DB 16; Length 822;						
Best Local Similarity 29.5%; Pred. No. 3.6e-44;						
Matches 272; Conservative						
Qy	1	SYELGLYQATVKNRVSVID---GKQATQKTENLTPEDEYSKREGINAEQIVIKITDQG	57			
Db	22	SYQLGKHMGGLATKDNQIAYIDDSKGKVPKPT-NKTMQDISAEGISAEQIVVVKITDQG	80			
Qy	58	YVTSHGSHYHYNGKVPDYAIISELLMKDNYLKDIEDIVNEKGGVYIVKDGKYYVYL	117			



```
Qy 756 GLRNNLTIQIMDNNSIMAEKLLALLKGSNPSVSKEK 794
Db 733 -----DKPKKQADK---SNENQOPSEASKEE 757

RESULT 13
Q8N282 PRELIMINARY; PRT; 823 AA.
ID Q8N282
AC Q8N282;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2003 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein spyM18_2072.
GN SPYM18_2072.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE010110; AAL98543.1; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 823 AA; 92585 MW; C79E1EB30CEBDF0C CRC64;

Query Match 22.7%; Score 946; DB 16; Length 823;
Best Local Similarity 29.9%; Pred. No. 1.1e-44;
Matches 261; Conservative 123; Mismatches 272; Indels 216; Gaps 29;

Qy 1 SYELGLQARTVKNNRVSVID---GKQATQKTENLTPDEVSKREGINAEQIVIKITDQG 57
Db 22 SYQLGKHHMGVPTKDNQIAYIDDSKGAKAPKT-NKTMWDQISAEQISAEQIVVKITDQG 80

Qy 58 YVTSHGDHYHYNGKVPYDAIISBEELMKDPNYKLKDEDIVNEKGGYVIVKDGKYYVYL 117
Db 81 YVTSHGDHYHYNGKVPYDAIISBEELMKDPNYRPFKQSDVINEILDGYVIVKVGNGYYVYL 140

Qy 118 KDAHADNVRTKEELNQKQHSQH-REGG-----TPRNDGVALARSQGRVTTDGY 169
Db 141 KPGSKRKNIRTKQIAEQVAKGTKEAKEKGLAQVAHLSKEEVAVNAEAKRQGRVTTDGY 200

Qy 170 IFNASDIETDGDYIVPHGPHYHYPKNELASAEALAAEFLS---GRGNLSRSRTYR 226
Db 201 IFSFTDIIDLDGAYLPHGNHYHYPKDLISPELAAQAYWQKQGRG--ARPSDYR 258

Qy 227 QNSNTSRTNWPVS-----NPGTNTNTSNTSNTNSQASQSDIDSLLK 272
Db 259 TPAQRRKAP-IPDVTNPNRQHQPNDGGYHPAPPNPNDASQNKHQDEFKGTFKELLD 317

Qy 273 QLYKLPLSORHVESDGLVDFDPAQITSRGAVVPHGPHYHYPYSQMSLEERARIIP 332
Db 318 QLRLDLKYRHEEDGLFEPTQVKSNAFYVVPHGPHYHIIIPRSQISPLEMELAD--- 374

Qy 333 LRYRSHNWVPSRPPSPQPTPEPSPGPQAPNLKIDNSLSVLSQVVRKVGEGVVFEEK 392
Db 375 -RYLAG-----QTEDNDGSDHDKPSDK-EVTHFFIGHRIKAYGKGLDGKPY 419

Qy 393 GIS-RYVPAKLPGETVKNLBSKLSQBSVSHLTAKENVAPRDQEFYDKAYNLLTEAH 451
Db 420 DTSDAYVFSK-----ESIHSV-----KSGVATAKGDHDFHYIGFGELEQYE 460
```







Qy	301	ARGVAVPHGDHYHFIPIYSQMSLEERETARIILPILRYNHNVDPDRPQSPQPTPEPSFG	360
Qy	302	PNGVAIPHGHDYHFIPIYSKLSALEEKIARWVP	333
Qy	361	PPAPNLKIDNSLSVQLRVKVGEGVVFEEKGISRVFAKDLPSFTVKNLESKLSQES	420
Db	334	-----ISGTGSTVSTNAK-----PNEVVSSLGSLSSNPSS	363
Qy	421	VSHTLTAKKNVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN	480
Db	364	----LTTSKELSSASDGVIFNPK-DIVEETATAYIVRHG--DHPHYIPK-----SNQIG	410
Qy	481	KEKLVDDLFLAPITHPERLGRKPNQSIETVEDVRIQLADKTTSDGYIFDEHDIISD	540
Db	411	OPTLPNNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFDA NR IIAE	452
Qy	541	EGDAYVPHMGHSHWICKDGLSDKEKVAQAAYTKEKGLPPSDADVKANPTGDSAAAIY	600
Db	453	DESGFVMSHGDNHNYFFKKDLTEEQIKAAQKGLKE-----VKTSHNGLDLSLSSH	501
Qy	601	NR-----VKGKRIPIPLRVLPYMY--BHTVEVKNGNLII--PHKDHYHINIIPAW	644
Db	502	EQDYPSNAKEMKDLDKKIEEKIAGIKQYGVKRESIVVNKKNXAIYIPHGDDHHA DP	558
Qy	645	FDQHTYKAPNGYTLDELFAITKYVVEHDPDRPHSNDG-----WGNASEHVLGKKDSDHSDPN	700
Db	559	IDEH---KPVGIG---HSHSNVELFPKEGVAKKGNKVYTGSELNVNVLKKNSTFNN	611
Qy	701	KNF 703	
Db	612	QNF 614	
RESULT	11		
Q8DQ07	ID	Q8DQ07 PRELIMINARY; PRT; 1039 AA.	
AC	Q8DQ07;		
DT	01-MAR-2003	(TRMBLrel. 23, Created)	
DT	01-MAR-2003	(TRMBLrel. 23, Last sequence update)	
DT	01-OCT-2003	(TRMBLrel. 25, Last annotation update)	
DE	Pneumococcal histidine triad protein E.		
GN	PHE OR SPR0908.		
OS	Streptococcus pneumoniae (strain ATCC BAA-255 / R6).		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=171101;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=21429245; PubMed=11544234;		
RA	Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,		
RA	DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geisinger C.,		
RA	Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,		
RA	LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,		
RA	McAhrn S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,		
RA	Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,		
RA	Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,		
RA	Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,		
RA	Glass J.J.;		
RT	"Genome of the bacterium Streptococcus pneumoniae strain R6.";		
RL	J. Bacteriol. 183:5709-5717(2001).		
DR	EMBL; AE008464; AAK9712.1; -.		
DR	PIR; D97985; D97985.		
DR	InterPro; IPR006270; Strep_his_triad.		
DR	Pfam; PF04270; strep_his_triad; 6.		
DR	TIGRFAMS; TIGR01363; strep_his_triad; 3.		
KW	Complete proteome.		
SQ	SEQUENCE 1039 AA; 114625 MW; 05CC226D2028F551 CRC64;		
Query Match	29.8%; Score 1243; DB 16; Length 1039;		
Best Local Similarity	41.2%; Pred. No. 3.7e-61;		
Matches 298; Conservative	94; Mismatches 180; Indels 152; Gaps 20;		

```

Db 141 AAHADNIRTKBEIKROKQSHSHMHGGGS--NDQAVVAARAQGRYTTDDGQYIFNASDIIED 198
QY 180 TGDAYIVPHGDHYHYIPKKNELASALAAAEFLSGRNLNSRTRYPRONSNDTSTRNWVP 239
Db 199 TGDAYIVPHGDHYHYIPKKNELASALAAAEFLSGRNLNSRTRYPRONSNDTSTRNWVP 247
QY 240 SVSNFGTNTNTSNNSTNSQASQSDNDISLLKQLKPLSORHVESDGLVDFPAQITSR 299
Db 248 NPAQPLSENHNLTVTPTVHQ--NOCENISLLRELKPLSERHVESDGLVDFPAQITSR 306
QY 300 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVDPSPPEQSPPTPEPSP 359
Db 307 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVDPSPPEQSPPTPEPSP 366
QY 360 GQOPAPNLK-IDSN---SSIVSOLVRKVGEGYVEEKGISRYVFAKDLSETVKNLESKL 415
Db 367 SQQAPNPQAPSNPIDEKLVKAVRKVGEGYVEEKGISRYVFAKDLSETVKNLESKL 426
QY 416 SKQESVSHLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
Db 427 AKQESLSHLKAKKTDLPSSDREFYKAYDILLARIHQDLLDNKGRQVDFEALDNLLERLK 486
QY 476 DESTNKEKLVDDLLAFIAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEH 535
Db 487 DVPSDKVLVDDILLAFIAPIRPERLGKFNQAITYTTDDBIQVAKLAGKYTTEDGYIFDPR 546
QY 536 DIISDEGDAYVTPHMGSHWIGKDSLDKEKVAQAAYTKKGLPPSPADVKANPTGDS 595
Db 547 DITSDEGDAYVTPHMGSHWIGKDSLDKEKVAQAAYTKKGLPPSPADVKANPTGDS 606
QY 596 AAAYNRVKGEKRIPLVRLPYMVVHTVEVKNGLIIPKHDXHNIKFAWDDHTYKAPNG 655
Db 607 AEALYNRVKAACKVPLDRMPYNLYQTVKNGSLIIPHYDHYNIKFEWDFEGLYEAPKG 666
QY 656 YLEDLFATIKYVVEHPDRPHSDNGWGNASHEVLGK-----KDHSE----- 697
Db 667 YLEDLFATIKYVVEHPDRPHSDNGWGNASHEVLGK-----KDHSE----- 726
QY 698 -----DPKNFKADEPEEETPAEPEVQVETKEVQAOLKAEVLL 738
Db 727 ESEKENHAGLPSADNLKPSDTEETEBAEDTTDEAEIPOVENSVINAKIADAEALL 786
QY 739 AKVTDSSLKANATETLAGLNNLTQIMDNNSIMAEKLLALLKGSNPSV 790
Db 787 EKVTPDSIQNAMEITLGLKSLLLGTKNNTISAEDVSLALLKESQAPI 838

RESULT 9
QBDQ08
ID Q8DQ08 PRELIMINARY; PRT; 853 AA.
AC Q8DQ08;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Pneumococcal histidine triad protein D.
GN PHTD OR SPR0907.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAlharen S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
```

```

RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99711.1; -.
DR PIR; C97985; C97985.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRPFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 853 AA; 95225 MW; 9BF06A1EDEB990CF5 CRC64;

Query Match 65.6%; Score 2733; DB 16; Length 853;
Best Local Similarity 63.5%; Pred. No. 1.1e-144;
Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;

QY 1 SYELGYYQARTV-KENNRVSYIDGKOATOKTENLTPDEVSKRGINAEQIVIKITDQGYV 59
Db 21 SYELGHRQAGQVKESNRVSYIDGQOQRAENLTPEVSKRGINAEQIVIKITDQGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISPELLMKDPNYQLKDSIDVNEIKGGYVVKVDGKYVYLKD 119
Db 81 TSHGDHYHYNGKVPYDAIISPELLMKDPNYQLKDSIDVNEIKGGYVVKVDGKYVYLKD 140
QY 120 AAHADNVRKKEINROKQESOHREGGTPRNDCAVALARSQRYTTDDGYIFNASDIIED 179
Db 141 AAHADNIRTKBEIKROKQESHSHNHS---RADNAVAARAQGRYTTDDGYIFNASDIIED 197
QY 180 TGDAYIVPHGDHYHYIPKKNELASALAAAEFLSGRNLNSRTRYPRONSNDTSTRNWVP 239
Db 198 TGDAYIVPHGDHYHYIPKKNELASALAAAEFLSGRNLNSRTRYPRONSNDTSTRNWVP 246
QY 240 SVSNFGTNTNTSNNSTNSQASQSDNDISLLKQLKPLSORHVESDGLVDFPAQITSR 299
Db 247 NPAQPLSENHNLTVTPTVHQ--NOCENISLLRELKPLSERHVESDGLVDFPAQITSR 305
QY 300 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVDPSPPEQSPPTPEPSP 359
Db 306 TARGVAVPHGDHYHYIPYQMSLEBRIARIIPLYRSNHWVDPSPPEQSPPTPEPSP 365
QY 360 GQOPAPNLK-IDSN---SSIVSOLVRKVGEGYVEEKGISRYVFAKDLSETVKNLESKL 415
Db 366 SQQAPNPQAPSNPIDEKLVKAVRKVGEGYVEEKGISRYVFAKDLSETVKNLESKL 425
QY 416 SKQESVSHLTAKENVAPRQDFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLN 475
Db 426 AKQESLSHLKAKKTDLPSSDREFYKAYDILLARIHQDLLDNKGRQVDFEALDNLLERLK 485
QY 476 DESTNKEKLVDDLLAFIAPITHPERLGKNSQIEYTEDVRIAQLADKYTTSDGYIFDEH 535
Db 486 DVPSDKVLVDDILLAFIAPIRPERLGKFNQAITYTTDDBIQVAKLAGKYTTEDGYIFDPR 545
QY 536 DIISDEGDAYVTPHMGSHWIGKDSLDKEKVAQAAYTKKGLPPSPADVKANPTGDS 595
Db 546 DITSDEGDAYVTPHMGSHWIGKDSLDKEKVAQAAYTKKGLPPSPADVKANPTGDS 605
QY 596 AAAYNRVKGEKRIPLVRLPYMVVHTVEVKNGLIIPKHDXHNIKFAWDDHTYKAPNG 655
Db 606 AEALYNRVKAACKVPLDRMPYNLYQTVKNGSLIIPHYDHYNIKFEWDFEGLYEAPKG 665
QY 656 YLEDLFATIKYVVEHPDRPHSDNGWGNASHEVLGK-----DHSEDPNK----- 701
Db 666 YLEDLFATIKYVVEHPDRPHSDNGWGNASHEVLGK-----DHSEDPNK----- 725
QY 702 -----NFKAD-----EPEVETPAEPEVQVE 723
Db 726 EEDKEHDEVSEPTHESEKENHVGNLPSADNLKPSDTEETEBAEDTTDEAEIPOVE 785
QY 724 TEKVBAQLKAEVLLAKVTDSSLKANATETLAGLNNLTQIMDNNSIMAEKLLALLK 783
Db 786 HSVINAKIAEAEALKEKVTDSIRQNAVETTLGLKSSLLLTGKNNNTISAEDVSLALLK 845
QY 784 GSNPSSV 790
Db 846 ESQPTPI 852
```

```
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Domitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RL proteins (the Pnt family) that are protective against sepsis.";
DR EMBL; AF318954; AAK06759.1; -
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Signal.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 819 819
SQ SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 66.9%; Score 2786.5; DB 2; Length 819;
Best Local Similarity 66.7%; Pred. No. 11e-147;
Matches 544; Conservative 84; Mismatches 140; Indels 47; Gaps 7;

Qy 1 SYELGLYQA-RTVKENNRVSVIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGYV 59
Db 21 SYELGRVQAGQDKESNRVAIDGDAQKAEINTPEVSKREGINAEQIVIKITDQGYV 80
Qy 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEIVNEVKGKVIKVDGKYVYLKD 119
Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSDIIVNEIKGKVIKVGKYYVYLKD 140
Qy 120 AAHADNVRTKEEINRQKQHSQHREGTTPRNDGVALARSQGRYTTDDGYIFNADIIED 179
Db 141 AAHADNVRTKEEIRKQKQHSQHNS---RADNVAARAAGRYTTDDGYIFNADIIED 197
Qy 180 TGDYIVPHGDHYHYIPKNSLSASELAEEAFLSGRNLNSRYSYRNSONTSTRTNWP 239
Db 198 TGDYIVPHGDHYHYIPKNSLSASELAEEAYWNG-----KQGRPSSSSSYNA 246
Qy 240 SVSNPGTNTNTNSNTNSQASQNDIDSLKQYKLPQSRHVESDGLVDPQAITSR 299
Db 247 NPAQRLSENHNLVTPTYHQ-NQGENISLLRELYAKPLSERHVESDGLVDPQAITSR 305
Qy 300 TARGVAVPHGDHYHFIYSQMSLEERITRIPIRYSNHWVPSRPEQSPQPTPEPSP 359
Db 306 TARGVAVPHGNHYHFIPEYQMSLEKRIARIPIRYSNHWVPSRPEPSPQPTPEPSP 365
Qy 360 GPQAPNLKIDNSNLSVQLVRKVGEGYVFEKGIYSYVFAKDLPSFTVKNLBSKSKQE 419
Db 366 SPQAPSNPID--GKLYKEAVRKVGDGVPFENGVSRYIPAKDLSAETAAGIDSKLAKQE 423
Qy 420 SVSHTLTAKENAVPRQDFVDKAYNLLTEAHKLFKNKGRNSQFQALDKLLERLNDEST 479
Db 424 SLSHKLGTKTKDPSDREFYKNAYDILLARIHQDLNKGQVDFEALDNLRLKDVSS 483
Qy 480 NKEKLVDDLLAFLAPITHPELGRKPNQIETEDVRLAQADKYTTSDDGYIFPEHIIIS 539
Db 484 DKVKLVDELLAFLAPIRHPERLGRKPNQITVTDDEIQVAKLAGKYTAEDGDYIFDPRIITS 543
Qy 540 DEGDAYVTPHMGSHWIKGDSLSDEKVAQAQYTKKGIPLPPSPDADVKNAPTGDSSAAAI 599
Db 544 DEGDAYVTPHMTSHWIKKDSLSSEARAAQAQYAXEKGLTPPSTDHQDSGNTAKGEAI 603
Qy 600 YNRVKGKRIPLVRLPVMVHTVVKGNLLIIPKHQVHNITKFAWFDHTYKAPNGYTL 659
Db 604 YNRVKAACKVPLDRMPYNLQYTVVEVKGSLIIPHYDHNLIKFEWFEGLYEAAPGYTL 663
Qy 660 DLFATIKYVEHPDRPHSDNGMGNASHVLGKDHSDPDKNFKADEE-----PVEET 713
Db 664 DLLATVIKYVEHPNERPHSONGFNASDHVQRNKGQADTNQTEKPEEKQTEKPEET 723
```

```
Qy 714 -----PABEVPQVETEKVEAQLKEAEVLLAKVTDSLSKANA 750
Db 724 PREEKPOSEKPEPKPTEPEESPEEPEVETEKVEEKLREAEEDLLGKTQDPIIKSNA 783
Qy 751 TETLAGLRNNLTLOIMDNNSIMAEAEKLLALLKGS 785
Db 784 KETLTGLKNLLFTGTDNNNTIMAEAEKLLALLKES 818

RESULT 8
Q9ANY2 PRELIMINARY; PRT; 839 AA.
ID Q9ANY2
AC Q9ANY2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHTD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Domitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RL proteins (the Pnt family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelini H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberger J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapflee E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AF318955; AAK06760.1; -
DR EMBL; AE007403; AAK75120.1; -
DR PIR; G95115; G95115.
DR TIGR; SP1003; -
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Signal, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 66.6%; Score 2772; DB 16; Length 839;
Best Local Similarity 65.0%; Pred. No. 7e-147;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

Qy 1 SYELGLYQARTV-KENNRVSVIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGYV 59
Db 21 SYELGRHQAGQVKESNRVSVIDGDAQKAEINTPEVSKREGINAEQIVIKITDQGYV 80
Qy 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEIVNEVKGKVIKVDGKYVYLKD 119
Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYQLKDSDIIVNEIKGKVIKVGKYYVYLKD 140
Qy 120 AAHADNVRTKEEINRQKQHSQHREGTTPRNDGVALARSQGRYTTDDGYIFNADIIED 179
```

QY 301 ARGVAVPHGDHYHFIPIYSOMSELEERARIIPLYRSNHWVPSRPEQSPQTPPSPG 360  
 Db 332 ANGVAVPHGDHYHFIPIYSQSPLEEKARIIPLYRSNHWVPSRPEQSPQTPPSPS 391  
 QY 361 POPAPNLK-IDSN---SSLSVLVRKVGEGYVFEKGISRYVFAKDLPSBTVKNLSKLS 416  
 Db 392 POPAPNPQAPSPNPIDEKLVKBAVRKVGEGYVFEENGVPRIIPAKDLSAETAAGIDSKLA 451  
 QY 417 QBSVSHTLTAKENVAPDRDQBYDYKAYNLLTEAHKALFNKGRNSDFQALDKLLERLND 476  
 Db 452 QBSLSHKLGAKTDLPSDDREFYNKAYDILLARIHQDLDNKGQVDFEALDNLLEKLD 511  
 QY 477 ESNTKEKLVDDLLAFAPITHPERLGPNSQIETVDEVRIAQLADKYTTSDGYIFDEHD 536  
 Db 512 VSSDKVLVDDLLAFAPITHPERLGPNAQIYTTDDEIQVAKLAGKYTTEDGYIFDPRD 571  
 QY 537 IISDEGDVYTPHMGSHWIGKSDSKKVAQAAYTKEKGIILPPSPDADVKANPTGDSA 596  
 Db 572 ITSDEGDVYTPHMTSHWIKKDSLSEARAAQAAYAKEKGLTPPSTDHQDSGNTAKGA 631  
 QY 597 AAIYNRVKGEKRIPLVRLPMVPEHTVEVKNGLIIPHKDHYHNIKFAWPDHHTYKAPNGY 656  
 Db 632 EAIYNRVKAACKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWFDGLYEAPKGY 691  
 QY 657 TLEDLPATIKYVVEHPDERPHSDGNGNASEHVLGKKHSDDPKNKFKADEE-----PV 710  
 Db 692 SLEDLLATKYVVEHPNERPHSDGNGNASHVQRNKGQADNTQTEKNEEKPOTEKPE 751  
 QY 711 EET-----PAPEVPQVETEKVAQLKEAEVLLAKVT 742  
 Db 752 EETPREKQSEKPEPKPTEPEESPESEEPQVETEKVKEKLEAEVLLAKVT 811  
 QY 743 DSSLKANAVETLAGLNLLTQIMDNNSIMAEKLLALLKGS 785  
 Db 812 NPIIKSNKATETLGLKNLLFGTDNNTIMAEKLLALLKES 854  
 RESULT 6  
 Q97QM9 PRELIMINARY; PRT; 819 AA.  
 ID Q97QM9  
 AC Q97QM9; 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Conserved domain protein.  
 GN SP1174.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]\_TaxID=1313;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC BAA-334 / TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapfle E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollinghead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 pneumoniae";  
 RL Science 293:498-506 (2001).  
 DR EMBL; AB007418; AAK75283.1; -.  
 DR PIR; B95136; B95136.  
 DR TIGR; SP1174; -.  
 DR InterPro; IPR006270; Strep\_his\_triad.  
 DR Pfam; PF04270; strep\_his\_triad; 5.  
 DR TIGRFAMS; TIGR01363; strep\_his\_triad; 2.  
 KW Complete proteome.

SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;  
 Query Match 67.2%; Score 2797.5; DB 16; Length 819;  
 Best Local Similarity 67.0%; Pred. No. 2.5e-148;  
 Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;  
 QY 1 SYELGLYQA-RTVKENNRVSYIDGKQATOKTENLTDEVSREGINAEQIVIKITDGYV 59  
 Db 21 SYELGRYQAGQDKKESNRVAYIDGQAGQKAENLTDEVSREGINAEQIVIKITDGYV 80  
 QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNPKLDEDIVNEVKGVIKVDGKYVYVLKD 119  
 Db 81 TSHGDHYHYNGKVPYDAIISELLMKDPNPKLDEDIVNEIKGGYVIVKNGYVYVLKD 140  
 QY 120 AAHADNVRTKEENRQKQHSQHSQREGTPRNDCAVALARSQRYTTDDGYIFNASDIIE 179  
 Db 141 AAHADNVRTKEENRQKQHSQHSQREGTPRNDCAVALARSQRYTTDDGYIFNASDIIE 197  
 QY 180 TGDYIVPHGDHYHYIPKNLSASLAAAFISGRGNLSNRTYRQNSDNTSRTNWVP 239  
 Db 198 TGDYIVPHGDHYHYIPKNLSASLAAAFISGRGNLSNRTYRQNSDNTSRTNWVP 246  
 QY 240 SVSNPGTNTNTNSNTNSQASQSDNDLSLLKQLYKLPLSRHVSDGLVFPDPAQITSR 299  
 Db 247 NPAQRLSENHNLTVPTVHQ-NQGENISLLRELYAKPLSERHVSDGLVFPDPAQITSR 305  
 QY 300 TAGVAVPHGDHYHFIPIYSOMSELEERARIIPLYRSNHWVPSRPEQSPQTPPSP 359  
 Db 306 TAGVAVPHGDHYHFIPIYSOMSELEERARIIPLYRSNHWVPSRPEQSPQTPPSP 365  
 QY 360 GPQAPNLKIDSNSLSVLVRKVGEGYVFEKGISRYVFAKDLPSBTVKNLSKLSKQE 419  
 Db 366 SPQAPNSNPID--EKLKVAVRKVGEGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQE 423  
 QY 420 SVSHHTTAKENVAPDRDQBYDYKAYNLLTEAHKALFNKGRNSDFQALDKLLERLND 479  
 Db 424 SLSHKLGTKTKTDLPSDDREFYNKAYDILLARIHQDLDNKGQVDFEALDNLLEKLDVSS 483  
 QY 480 NKEKLVDDLLAFAPITHPERLGPNSQIETVDEVRIAQLADKYTTSDGYIFDEHDIIIS 539  
 Db 484 DKVKLVDDLLAFAPITHPERLGPNAQIYTTDDEIQVAKLAGKYTTEDGYIFDPRIITS 543  
 QY 540 DEGDVYTPHMGSHWIGKSDSKKVAQAAYTKEKGIILPPSPDADVKANPTGDSAAA 599  
 Db 544 DEGDVYTPHMTSHWIKKDSLSEARAAQAAYAKEKGLTPPSTDHQDSGNTAKGAEL 603  
 QY 600 YNRVKEKRIPLVRLPMVPEHTVEVKNGLIIPHKDHYHNIKFAWPDHHTYKAPNGYTL 659  
 Db 604 YNRVKAACKVPLDRMPYNLQYTVVEKNGSLIIPHYDHYHNIKFEWFDGLYEAPKGYTL 663  
 QY 660 DLFPATIKYVVEHPDERPHSDGNGNASEHVLGKKHSDDPKNKFKADEE-----PVEET 713  
 Db 664 DLLATKYVVEHPNERPHSDGNGNASHVQRNKGQADNTQTEKSEEPQOTEKPEET 723  
 QY 714 -----PABPEVPQVETEKVAQLKEAEVLLAKVTDSLSKANA 750  
 Db 724 PREKQSEKPEPKPTEPEESPESEEPQVETEKVKEKLEAEVLLAKVTDSLSKANA 783  
 QY 751 TETLAGLNLLTQIMDNNSIMAEKLLALLKGS 785  
 Db 784 KETLTGLKNLLFGTDNNTIMAEKLLALLKES 818

RESULT 7  
 Q9ANY3 PRELIMINARY; PRT; 819 AA.  
 ID Q9ANY3  
 AC Q9ANY3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Pneumococcal histidine triad protein B precursor (Fragment).  
 GN PHTB.  
 OS Streptococcus pneumoniae.



QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180  
Db 141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 200  
QY 181 GDAYIVPHGDHGHYIIPKNELASLAAAFISGRGNLSNRSRYRQNSDNTSRTNWVPS 240  
Db 201 GDAYIVPHGDHGHYIIPKNELASLAAAFISGRGNLSNRSRYRQNSDNTSRTNWVPS 260  
QY 241 VSNPGTNTNTSNNSTNSQASQNSDIDSLKQYKLPISQRHVSDGLVDFDPAQITST 300  
Db 261 VSNPGTNTNTSNNSTNSQASQNSDIDSLKQYKLPISQRHVSDGLVDFDPAQITST 320  
QY 301 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHNWVPSRPEQSPQPTPEPSPG 360  
Db 321 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHNWVPSRPEQSPQPTPEPSPG 380  
QY 361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSKQES 420  
Db 381 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSKQES 440  
QY 421 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480  
Db 441 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500  
QY 481 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540  
Db 501 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560  
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAIY 600  
Db 561 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAIY 620  
QY 601 NRVGKGRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHDTYKAPNGYTTLED 660  
Db 621 NRVGKGRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHDTYKAPNGYTTLED 680  
QY 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPVEETPAPEPVP 720  
Db 681 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPVEETPAPEPVP 740  
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 780  
Db 741 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 800  
QY 781 LLKGSNPSSVSKEKIN 796  
Db 801 LLKGSNPSSVSKEKIN 816

## RESULT 3

## Q8DPQ2

ID Q8DPQ2 PRELIMINARY; PRT; 828 AA.  
AC Q8DPQ2;  
DT 01-MAR-2003 (TEMBLrel. 23, Created)  
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Pneumococcal histidine triad protein A.  
GN PHTA OR SPRL061.  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21429245; PubMed=11544234;  
RA Hoskins J., Alborn W.B. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,  
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,

RA Glass J.I.;  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
RL J. Bacteriol. 183:5709-5717(2001).  
DR EMBL; AB008479; AK99865.1; -.  
DR PIR; E98004; E98004.  
DR InterPro; IPR006270; Strep\_his\_triad.  
DR Pfam; PF04270; strep\_his\_triad; 5.  
DR TIGRFAMs; TIGR01363; strep\_his\_triad; 2.  
KW Complete proteome.  
SQ SEQUENCE 828 AA; 93015 MW; 12CCCF407B550CID CRC64;

Query Match 99.9%; Score 4159; DB 16; Length 828;  
Best Local Similarity 99.7%; Pred. No. 1.6e-224;  
Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYELGLYQARTKVENNRVSYIDGKQATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 60  
Db 33 SYELGLYQARTKVENNRVSYIDGKQATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 92  
QY 61 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLDKEDI VNEVKGYYIKVDGKYYVYLKDA 120  
Db 93 SHGDHYHYNGKVPYDAIISSELLMKDPNYKLDKEDI VNEVKGYYIKVDGKYYVYLKDA 152  
QY 121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180  
Db 153 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 212  
QY 181 GDAYIVPHGDHGHYIIPKNELASLAAAFISGRGNLSNRSRYRQNSDNTSRTNWVPS 240  
Db 213 GDAYIVPHGDHGHYIIPKNELASLAAAFISGRGNLSNRSRYRQNSDNTSRTNWVPS 272  
QY 241 VSNPGTNTNTSNNSTNSQASQNSDIDSLKQYKLPISQRHVSDGLVDFDPAQITST 300  
Db 273 VSNPGTNTNTSNNSTNSQASQNSDIDSLKQYKLPISQRHVSDGLVDFDPAQITST 332  
QY 301 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHNWVPSRPEQSPQPTPEPSPG 360  
Db 333 ARGVAVPHGDHGHYIIPYSQMSLEERARIIPLYRSNHNWVPSRPEQSPQPTPEPSPG 392  
QY 361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSKQES 420  
Db 393 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSRTVKNLESKLSKQES 452  
QY 421 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480  
Db 453 VSHTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 512  
QY 481 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540  
Db 513 KEKLVDDLLAFLAPIITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 572  
QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAIY 600  
Db 573 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTKKGIILPPSPDADVKANPTGDSAAIY 632  
QY 601 NRVGKGRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHDTYKAPNGYTTLED 660  
Db 633 NRVGKGRIPLVRLPYMVEHTVEVKNGLIIPKHQHYHNIKFAWFDHDTYKAPNGYTTLED 692  
QY 661 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPVEETPAPEPVP 720  
Db 693 LFATIKYVVEHPDERPHSDNGWGNASEHVLGKDHSEDPNKNPKADEEPVEETPAPEPVP 752  
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 780  
Db 753 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLA 812  
QY 781 LLKGSNPSSVSKEKIN 796  
Db 813 LLKGSNPSSVSKEKIN 828

## RESULT 4

```

90 155 3.7 869 4 Q9NYF8 Q9nyf8 homo sapien
ALIGNMENTS
ID Q97QM8 PRELIMINARY; PRT; 802 AA.
AC Q97QM8;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Conserved domain protein.
GN SP1175.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AB007418; AAK75284.1; -.
DR PIR; C95136; C95136.
DR TIGR; SP1175; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4F5CB8364EEA1833 CRC64;

Query Match 100.0%; Score 4163; DB 16; Length 802;
Best Local Similarity 99.9%; Pred. No. 9e-225;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 7 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 66
Qy 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEKGGYVVKDGGYVYLKDA 120
Db 67 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEKGGYVVKDGGYVYLKDA 126
Qy 121 AHADNVRTKEINQKQHSQHRREGGTPRNDGAVALARQSGRYTTDDGYIENASDIIDT 180
Db 127 AHADNVRTKEINQKQHSQHRREGGTPRNDGAVALARQSGRYTTDDGYIENASDIIDT 186
Qy 181 GDAYIVPHGDHYHYIPKNELSASELAAAEAFLSGRGNLSNRTYRRQNSDTSRTNWVPS 240
Db 187 GDAYIVPHGDHYHYIPKNELSASELAAAEAFLSGRGNLSNRTYRRQNSDTSRTNWVPS 246
Qy 241 VSNPCTTNTNTSNNTNTSQASQNDIDSLKQLYKPLSORHVESDGLVDPDAQITSR 300
Db 247 VSNPCTTNTNTSNNTNTSQASQNDIDSLKQLYKPLSORHVESDGLVDPDAQITSR 306
Qy 301 ARGVAVPHGDHYHYIPYSQMSLEERIIARIIPLYRSNHWVPDSRPEQSPQPTPEPSG 360
Db 307 ARGVAVPHGDHYHYIPYSQMSLEERIIARIIPLYRSNHWVPDSRPEQSPQPTPEPSG 366
Qy 361 PQAPNLKIDNSSLSVLQVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQES 420

```

```

Db 367 PQAPNLKIDNSSLSVLQVRKVGEGYVFEKGISRYVFAKDLPSFTVKNLESKLSKQES 426
Qy 421 VSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGKNSDFQALDKLLERLNDESTN 480
Db 427 VSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDESTN 486
Qy 481 KEKLVDLLAFLAPITTHPERLGKPNQSIETEVTEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db 487 KEKLVDLLAFLAPITTHPERLGKPNQSIETEVTEDEVRIAQLADKYTTSDGYIFDEHDIISD 546
Qy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQAYTKEKGLTPSPDADVKANPTGSAAYI 600
Db 547 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQAYTKEKGLTPSPDADVKANPTGSAAYI 606
Qy 601 NRKGEKRIPLVRLPYMVEHTVEVKNGLNLIIPHDKHYHNIKPAWFDHDTYKAPNGYTTLED 660
Db 607 NRKGEKRIPLVRLPYMVEHTVEVKNGLNLIIPHDKHYHNIKPAWFDHDTYKAPNGYTTLED 666
Qy 661 LFATIKYVVEHPDERPHSDNGWGNASBHVLGKKDHSDDPNKFKADEPVEETPAEPVP 720
Db 667 LFATIKYVVEHPDERPHSDNGWGNASBHVLGKKDHSDDPNKFKADEPVEETPAEPVP 726
Qy 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTQIMDNNSIMAEAKLLA 780
Db 727 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTQIMDNNSIMAEAKLLA 786
Qy 781 LLKGSNPFSSVSKEKIN 796
Db 787 LLKGSNPFSSVSKEKIN 802

RESULT 2
Q9AHT9 PRELIMINARY; PRT; 816 AA.
ID Q9AHT9;
AC Q9AHT9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
RT Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL; AF291695; AAK19155.1; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 100.0%; Score 4163; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 9.3e-225;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 60
Db 21 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVT 80
Qy 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEKGGYVVKDGGYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEKGGYVVKDGGYVYLKDA 140

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 05:49:44 ; Search time 77 Seconds

(without alignments)  
3261.721 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTYKENRVSY.....KLLALLKGNPSSVSKEKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

1: SPTRMBL 25:\*

2: sp\_archaea:\*

3: sp\_bacteria:\*

4: sp\_fungi:\*

5: sp\_human:\*

6: sp\_invertebrate:\*

7: sp\_mammal:\*

8: sp\_mhc:\*

9: sp\_organelle:\*

10: sp\_phase:\*

11: sp\_plant:\*

12: sp\_rodent:\*

13: sp\_virus:\*

14: sp\_vertebrate:\*

15: sp\_unclassified:\*

16: sp\_rvirus:\*

17: sp\_bacteriaph:\*

18: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	802	16	Q97QM8 streptococc
2	4163	100.0	816	2	Q9AHT9 streptococc
3	4159	99.9	828	16	Q8DPQ2 streptococc
4	3236	77.7	844	2	Q9AG74 streptococc
5	3224	77.4	855	16	Q8CWR4 streptococc
6	2797.5	67.2	819	16	Q97QM9 streptococc
7	2786.5	66.9	819	2	Q9ANY3 streptococc
8	2772	66.6	839	16	Q9ANY2 streptococc
9	2733	65.6	853	16	Q8DQ08 streptococc
10	1246	29.9	1039	16	Q9ANY1 streptococc
11	1243	29.8	1039	16	Q8DQ07 streptococc
12	949	22.8	825	16	Q99XV4 streptococc
13	946	22.7	823	16	Q8NZ82 streptococc
14	945	22.7	822	16	Q8E4U1 streptococc
15	944	22.7	825	2	Q93GT5 streptococc
16	940	22.6	823	16	Q8K5Q1 streptococc

Q8dz81 streptococc	17	937	22.5	822	16	Q8DZ81
Q9zhg7 streptococc	18	934	22.4	822	2	Q9ZHG7
Q877Y2 streptococc	19	922	22.1	794	16	Q877Y2
Q8e338 streptococc	20	916.5	22.0	481	16	Q8E338
Q9ae1 streptococc	21	679.5	16.3	289	2	Q9AE21
Q8esr2 streptococc	22	305	7.3	877	16	Q8ESR2
Q8e029 streptococc	23	302	7.3	877	16	Q8E029
Q8e029 streptococc	24	243	5.8	182	16	Q8DQ06
Q99276 streptococc	25	223	5.4	792	16	Q99276
Q8k714 streptococc	26	213	5.1	792	16	Q8K714
Q963t1 plasmodium	27	197.5	4.7	1078	5	Q963T1
Q77033 dictyosteli	28	192.5	4.6	1390	5	Q77033
Q8p0g5 streptococc	29	192	4.6	792	16	Q8P0G5
Q8ict8 plasmodium	30	181.5	4.4	2752	5	Q8ICT8
Q15754 dictyosteli	31	178	4.3	565	5	Q15754
Q8i56 plasmodium	32	178	4.3	1233	5	Q8I56
Q8ib94 plasmodium	33	177	4.2	8591	5	Q8IB94
Q25860 plasmodium	34	176.5	4.2	1271	5	Q25860
Q9gtx2 plasmodium	35	175.5	4.2	1236	5	Q9GTX2
Q8ihn3 plasmodium	36	175.5	4.2	5507	5	Q8IHN3
Q8icv5 plasmodium	37	173	4.2	1063	5	Q8ICV5
Q9u459 plasmodium	38	172.5	4.1	5458	5	Q9U459
Q9lci9 staphylococ	39	172	4.1	1795	16	Q9LCJ9
Q9icn2 staphylococ	40	172	4.1	2478	2	Q9ICN2
Q99gr6 staphylococ	41	172	4.1	2481	16	Q99GR6
Q8iea3 plasmodium	42	171.5	4.1	3452	5	Q8IEA3
Q9rl69 staphylococ	43	170	4.1	2478	2	Q9RL69
Q82345 arabidopsis	44	168.5	4.0	1043	10	Q82345
Q8i413 plasmodium	45	168.5	4.0	1708	5	Q8I413
Q7x0p4 lactobacill	46	167.5	4.0	4427	2	Q7X0P4
Q8ieb6 plasmodium	47	166.5	4.0	2651	5	Q8IEB6
Q8i223 plasmodium	48	166	4.0	1465	5	Q8I223
Q97255 plasmodium	49	166	4.0	1826	5	Q97255
Q9nfs3 drosophila	50	166	4.0	16215	5	Q9NFS3
Q9i7u4 drosophila	51	166	4.0	18074	5	Q9I7U4
Q8i3a8 plasmodium	52	165.5	4.0	1373	5	Q8I3A8
Q8i121 plasmodium	53	165.5	4.0	2227	5	Q8I121
Q8i2v4 plasmodium	54	165.5	4.0	3381	5	Q8I2V4
Q99td3 staphylococ	55	165	4.0	891	16	Q99TD3
Q8ze8 lactobacill	56	165	4.0	1231	16	Q8ZE8
Q8i1n9 plasmodium	57	165	4.0	3370	5	Q8I1N9
Q8i1p4 staphylococ	58	164	3.9	891	16	Q8I1P4
Q8idp7 plasmodium	59	162.5	3.9	1488	5	Q8IDP7
Q9u4x0 plasmodium	60	162.5	3.9	2647	5	Q9U4X0
Q8nw39 staphylococ	61	162	3.9	895	16	Q8NW39
Q7z3u2 homo sapien	62	162	3.9	1118	4	Q7Z3U2
Q8i1d3 plasmodium	63	162	3.9	2467	5	Q8I1D3
Q9fnd5 arabidopsis	64	162	3.9	2910	10	Q9FND5
Q8i145 plasmodium	65	162	3.9	3504	5	Q8I145
Q8i1l2 plasmodium	66	162	3.9	4405	5	Q8I1L2
Q9vp19 drosophila	67	162	3.9	5322	5	Q9VPL9
Q8i659 plasmodium	68	161.5	3.9	1383	5	Q8I659
Q8ibl1 plasmodium	69	160.5	3.9	1003	5	Q8IBL1
Q8i5h0 plasmodium	70	160.5	3.9	1461	5	Q8I5H0
Q8cnu9 staphylococ	71	160	3.8	3692	16	Q8CNU9
Q8x1l1 clostridium	72	159.5	3.8	1129	16	Q8X1L1
Q8cp76 staphylococ	73	159	3.8	9439	16	Q8CP76
Q83k2 enterococcu	74	158.5	3.8	1499	16	Q83K2
Q77328 plasmodium	75	158	3.8	1650	5	Q77328
Q8ied3 plasmodium	76	158	3.8	1840	5	Q8IED3
Q8ic27 plasmodium	77	157	3.8	1859	5	Q8IC27
Q8i1u7 plasmodium	78	156.5	3.8	1139	5	Q8I1U7
Q8h6x1 arabidopsis	79	156.5	3.8	1157	10	Q8H6X1
Q8ime5 drosophila	80	156.5	3.8	1185	5	Q8IME5
Q9lza8 arabidopsis	81	156.5	3.8	1495	10	Q9LZA8
Q9s1g8 arabidopsis	82	156	3.7	519	10	Q9S1G8
Q8idb6 plasmodium	83	156	3.7	1692	5	Q8IDB6
Q7za38 ashba goss	84	156	3.7	3392	3	Q7ZA38
Q9p619 schizosacch	85	155.5	3.7	1038	3	Q9P6L9
Q9veu8 drosophila	86	155.5	3.7	1298	5	Q9VEU8
Q8cgb3 mus musculu	87	155.5	3.7	1413	11	Q8CGB3
Q8i350 plasmodium	88	155.5	3.7	3069	5	Q8I350
Q86wu6 homo sapien	89	155	3.7	689	4	Q86WU6





240	QY	SVYNGPCT-----NTNTSNNSTNSQAS-QSNDIDSLAKQLYKPLSQRHVES	286
1151	Db	ASDPPASSEMSPSTOLLFYETSAGFSEVLVLQPSQASDVTLTKTV--LPA-----VPS	1204
287	QY	DGLVFPDAQ-----ITSRTARGVAVPHGDHYHFIPYSOMSELEERARIPIURY	335
1205	Db	DFILVETPKVDKLSSTMLHLIIVSNSASSEMMLHSTSPVFDVSPSTHMHASLQIGITISY	1264
336	QY	RGNHWVPDRPEQSPQPTPEPSGPQOPAPNLIKDSNLSVLQVRKVGQGVVFPEKGIS	395
1265	Db	ASEKYEPVLLKSESSHQVVP-----SLYSNDELFTQANLEINQAH--PPKG--	1308
396	QY	RYVPFADKLPB--ETVKNLSEKLSQBSVSHTLTAKENVAPR-----DOEFYDKAY	444
1309	Db	RHYFATPVLISIDPLNTLINKLHSDIEI---LFTSKSVTGKVFAGIPTVASDTFVSTDH	1365
445	QY	NL--LTEAH-----KALFXNKG-----	459
1366	Db	SVPIGNGHVAITAVSPHRDGSVTSKLLFPSPKATSELSHSAKSDAGLVGGEDGDTDDG	1425
460	QY	-----RNSDFQALDKLLERLNDESTNKEKLVDDLLAFIPIHPERLGPKNQIEV----	510
1426	Db	DDDDDRSDGSLTHKMS--CSSYRESQEKVMND-----SDTHENSIMDQNNPIYSLSLE	1478
511	QY	TEDEVRIAGLADKYTT-----SDGYFDFEHI-----ISDRG	542
1479	Db	NSEEDNRKTVSSDSQTDGMRSPCKSPSANGLSQKINDGK--EENDIQTGSALLPLSPES	1536
543	QY	DAYVTPHMGSHWIGK--DLSLDEKVA--AAQYTKEX--GIL-----PPSP	583
1537	Db	KAAVLTSDDESGGQGTSDSLNENETSDFSPADTNEKDADGILAAAGDSEITPGFPQSP	1596
584	QY	DADV-----KANPTGDSAAAIYNEVKGEKR--IPLVRLPYMVHEHTVVKXG	627
1597	Db	TSSVTSENSEVFHVSEABASNSHSEGRIGLAEGSEKKAVIDPLIVSALFTFCLVLVVG	1656
628	QY	NL-----IIPKDHVHNIRKFAWFDHHT--YKAPNGY	656
1657	Db	ILIYWRKCFQTAHFYLEDSTSPRVISPTPTPIFISDDVGAIPIKHFKPHVADLHASSGF	1716
657	QY	TLDELPAATIK-YVVE-----HPPERPHSN	679
1717	Db	TEE--FETLKEFYQEVSQCTVDLIGTADSSNHFDNK-HKN	1753

Search completed: October 1, 2004, 07:15:18  
Job time : 34 secs

"The cloning of a receptor-type protein tyrosine phosphatase expressed in the central nervous system.";  
 J. Biol. Chem. 268:10573-10581(1993).  
 [3]  
 RT SEQUENCE OF 1479-2091 FROM N.A. (ISOFORM LONG).  
 RP TISSUE=Liver;  
 RX MEDLINE=91006018; PubMed=2170109;  
 RA Krueger N.X., Streuli M., Saito H.;  
 RT "Structural diversity and evolution of human receptor-like protein tyrosine phosphatases.";  
 RL EMBO J. 9:3241-3252(1990).  
 [4]  
 RP SEQUENCE OF 1749-1990 AND 2047-2280 FROM N.A. (ISOFORM LONG).  
 RX TISSUE=Brain stem; PubMed=2169617;  
 RA Kaplan R., Morse B., Huebner K., Croce C., Howk R., Ravera M., Ricca G., Jaye M., Schlössinger J.;  
 RT "Cloning of three human tyrosine phosphatases reveals a multigene family of receptor-linked protein-tyrosine-phosphatases expressed in brain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).  
 CC -!- FUNCTION: May be involved in the regulation of specific developmental processes in the CNS.  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.  
 CC -!- SUBUNIT: The carbonic-anhydrase like domain binds to contactin (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P23471-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P23471-2; Sequence=VSP\_005151;  
 CC -!- TISSUE SPECIFICITY: Specifically expressed in the central nervous system, where it is localized in the Purkinje cell layer of the cerebellum, the dentate gyrus, and the subependymal layer of the anterior horn of the lateral ventricle. Developmentally regulated in the brain.  
 CC -!- SIMILARITY: Contains 1 eukaryotic-type carbonic anhydrase domain.  
 CC -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -!- CAUTION: Called RPPase beta in Ref.2 and Ref.4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M93426; AAA60225.1; -;  
 CC EMBL; X54135; CAA38070.1; -;  
 CC PIR; A46151; A46151.  
 CC HSP; P18052; 1YFO.  
 CC Genew; HGNC:9685; PTPR21.  
 CC MIM; 176891; -;  
 CC GO; GO:0005887; C:integral to plasma membrane; TAS.  
 CC GO; GO:0005001; F:transmembrane receptor protein tyrosine phosphatase; TAS.  
 CC GO; GO:0007417; P:central nervous system development; TAS.  
 CC GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.  
 CC InterPro; IPR001148; Euk\_Coahnd.  
 CC InterPro; IPR008957; FN\_III-like.  
 CC InterPro; IPR003961; FN\_III.  
 CC InterPro; IPR000387; TYR\_phosphatase.  
 CC InterPro; IPR000242; Tyr\_PP.  
 CC Pfam; PF00194; carb\_anhydrase; 1.  
 CC Pfam; PF00041; fn3\_1.  
 CC Pfam; PF00102; Y\_phosphatase; 2.  
 CC PRINTS; PR00700; PRTYPHPHASE.  
 CC ProDom; PD000865; Euk\_Coahnd; 1.  
 CC SMART; SM00060; FN3; 1.

DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 2.  
 KW Signal; glycoprotein; Transmembrane; Hydrolase; Repeat;  
 KW Alternative splicing.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 2314 RECEPTOR-TYPE PROTEIN-TYROSINE  
 FT PHOSPHATASE ZETA.  
 FT POTENTIAL.  
 FT DOMAIN 25 1635 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 1636 1661 CARBONIC-ANHYDRASE LIKE.  
 FT DOMAIN 1662 2314 FIBRONECTIN TYPE-III.  
 FT DOMAIN 34 302 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 312 406 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT DOMAIN 1744 1997 PHOSPHOCYSTEINE INTERMEDIATE (BY  
 FT ACT\_SITE 1932 1932 SIMILARITY).  
 FT SITE 2222 2222 ANCESTRAL ACTIVE SITE.  
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 587 587 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)  
 FT (POTENTIAL).  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 637 637 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)  
 FT (POTENTIAL).  
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 997 997 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)  
 FT (POTENTIAL).  
 FT CARBOHYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1122 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1456 1456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1548 1548 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)  
 FT (POTENTIAL).  
 FT CARBOHYD 1550 1550 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)  
 FT (POTENTIAL).  
 FT CARBOHYD 1561 1561 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1617 1617 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 755 1614 Missing (in isoform Short).  
 FT /FTId=VSP\_005151.2).  
 FT CONFLICT 1722 1728 MISSING (IN REF. 2).  
 FT SQ SEQUENCE 2314 AA; 254528 MW; 77DBDEF4A0F5FB42 CRC64;  
 Query Match 3.4%; Score 141.5; DB 1; Length 2314;  
 Best Local Similarity 19.2%; Pred. No. 13;  
 Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 43;  
 QY 51 IKITDQGYVTHSGDHYHYNGKVPYDAIISSELLMKDPNPKLXDEIVNEVGKGYIKVD 110  
 Db 954 VGVTYQGSILFSGPSHI-----PIPKSLITPTASLLQPTHALSG-----D 993  
 QY 111 GKYYVYLKQAAHADNVRTKEINRQKQHSQHREGGTGRNDGAVAL-----ARSGRYT 164  
 Db 994 GEW-----SGASDS-----EFLLPDTDLTALNISPSVABFTY 1030  
 QY 165 T-----DDGYIFNADIIEDTGDAYIVPHGDHYHY-----IPK-----NELSASELAAAE 209  
 Db 1031 TSVFGDDNKALKSKSELIYGNTELOIPSNEMVYPSSEIVPMYDYNVKNLASLETSTV 1090  
 QY 210 AFLSGRG-----NLNSRTYR-RQNSDNTS-----RTNWVP 239  
 Db 1091 SISSTKGMFPFSLAHTTTTKVFDHEISQVPENNFVQPTHVTSQASGDTSLKPLVLSANSEP 1150





RL Neuron 9:417-428 (1992).  
 CC -!- TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.  
 CC -!- DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURULA AND PERSIST  
 CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL  
 CC CORD.  
 CC -!- SIMILARITY: Belongs to the intermediate filament family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M99387; AAA49966.1; -;  
 CC PIR; JH0720; JH0720.  
 CC InterPro; IPR001664; IF.  
 CC Pfam; PF00038; filament; 1.  
 CC PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone.  
 FT DOMAIN 1 12  
 FT HEAD.  
 FT DOMAIN 13 314  
 FT ROD.  
 FT DOMAIN 315 1744  
 FT TAIL.  
 FT DOMAIN 8 48  
 FT COIL 1A.  
 FT DOMAIN 49 60  
 FT LINKER 1.  
 FT DOMAIN 61 156  
 FT COIL 1B.  
 FT DOMAIN 157 179  
 FT LINKER 12.  
 FT DOMAIN 180 193  
 FT COIL 2A.  
 FT DOMAIN 194 199  
 FT LINKER 2.  
 FT DOMAIN 200 314  
 FT COIL 2B.  
 SQ SEQUENCE 1744 AA; 199561 MW; 6502EAC9F6C4E93 CRC64;  
 Query Match 3.4%; Score 142; DB 1; Length 1744;  
 Best Local Similarity 17.1%; Pred. No. 8.3;  
 Matches 144; Conservative 145; Mismatches 308; Indels 244; Gaps 37;  
 QY 8 QARTKNNRVSYDGKQATKNTLTPDEVSKR-----EGINAE-QIVIKI---TDQ 56  
 DB 659 KKKVPLENEYIPVSKDDLTETSHLENDSESSQSFDSKLFENKSTEDQLITNLKNTQE 718  
 QY 57 GYVTSHGHDHYHYNGKVPYDAI-----SELLMKDPN-----YKLKBDIVN 99  
 DB 719 NIFQSNQEHLE---NLEFDSVVPDVKFMYQENLLBEENVYGDGLVQMATDENIIN 774  
 QY 100 EVKGGVVIKGVYVYLKDAHADNVRTKEI---NRQKQHSQHREGGTPRNDGAVA 155  
 DB 775 Q-----SSDQLLLSDHSHHEETKTSSTAVEHNRMESEHAE-----VDKSS 816  
 QY 156 LARSQRYTTDDGYIFNASDITETGDYIVPHGDHYHVPKNELSASELAFAFLSGR 215  
 DB 817 IPVEISENVSVREIHEISDVEDTKQAF----- 845  
 QY 216 GNLSNRTYRRQNSDNTSTNNVPSVSNPQTNTNTSNNSTNSQASQNSDLSLLKQLY 275  
 DB 846 ---EDERVEQINQNNQEST-----VDLDGSVYSQENSQLEDEVSIEQIEKDFE--- 894  
 QY 276 KLPLSQRHVESDGL--VFDPQAITSRTAGVAPHGDHYHFFPYQSMELEERARIPL 333  
 DB 895 ---INEQELKSKQIRAEAFTEEV-----DH-QVDFMQEQSFEREVGQLNNI 938  
 QY 334 R-----YRSNHWVPSRPEQSPQTPPEPSFGPOPAPNLKIDGNSLSVQLVRKVGEGYVF 389  
 DB 939 KQEVYLYQNY-----DEDSFQNNDEPQ-----ELES-CDLQEQKIKLEENQLS 981  
 QY 390 EKGISRY-----VPFAK-----DLPSVTKNML-----ESKLSQESVSHTL 425  
 DB 982 ENENQNFNGNDIEEFSQGYDTEICQETIGNQVSAQLLCESDINQDKLSMEDEEQNN 1041  
 QY 426 TAKKENVAPR-----DQFYDKAYNLLTEAHKALFKNKGNSDFDQALD- 468  
 DB 1042 PETEDNIGLQESQDNTSRNEGTFKFSQECVDFVFKPDMDSKSEYSGQOEDLDKQVTF 1101

QY 469 -----KLLER-----LNDESTNKEKLVDDL--AFLAPITHPERLCKPMS 506  
 DB 1102 SLNEQANDLLEKEEVILHHADDQRSVNDIITDEKLSERIIDNELATVDVNESLAANKE 1161  
 QY 507 QLEYTEDEVRIAQL-----ADKYTTSDGYIFDEHDIISDEGDAYVTHMGSHWIGKD 559  
 DB 1162 QVDLFTDEYAVDNDVGMQDDSGQYQTKEDLFDVGNII-----EKIIOQT 1208  
 QY 560 SLSDKEKVAQAAYTKEGILPPSPDADVKANPTGDSAAAIYNRVKGEKRI-PLVRLPYMV 618  
 DB 1209 SILNQB-ICERVNDVDEDSIGAKNESVEMNDVDLVEPA--KVTGDEQISPLQDEKLN 1265  
 QY 619 EHTVEVK--NGNLIIPHDKDHYHNKIPA-----WFDHHTYKAPNGYTLDELFAIKYVVEHP 672  
 DB 1266 ETMEDTKMDGQCLCEKENETEIEVTDSQPQATDLSHDAGRELTVDSQSANLIQ-FCENP 1324  
 QY 673 DERPHSNDGNGNASEHVLGKDHSDPDNKNFKADEEPVEET---PAEPEVPQVETEKVEA 729  
 DB 1325 TKT-----LIAHHI-----EYETVADSDLESTEEQVQETERIPPKPDSKMNENSES 1372  
 QY 730 Q 730  
 DB 1373 E 1373  
 RESULT 23  
 MOT3 YEAST STANDARD; PRT; 490 AA.  
 ID MOT3 YEAST  
 AC P54785; 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein MOT3/HMS1.  
 GN MOT3 OR HMS1 OR YMR070W OR YMR916.09.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RA Madison J., Winston F.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=97313268; PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT XIII.";  
 RL Nature 387:90-93 (1997).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U25279; AAC49982.1; -;  
 CC EMBL; Z48952; CA88795.1; -;  
 CC PIR; S52830; S52830.  
 CC GeneOnline; 142736; -;  
 CC TRANSFAC; T03448; -;  
 CC TRANSFAC; T03500; -;  
 CC SGD; S0004674; MOT3.  
 CC GO; GO:0005634; C:nucleus; IDA.

Db 212 RNLDQWRPSSLSQTSNTPHAANPSFPGTIVHTNFRPREGGCHRRRSTGSL----SVG 268

QY 382 KVGEVYFEEKISR-YVFAKDLPSFTVKNLESKLSQESVSHLTAKKNVAPRQDEFY 440

Db 269 SSGSGFSSGSGNPRKNLSPYLPQSSIPAL---LAERRLVTGILIVSKKN---RSDAFV 322

QY 441 DRAYNLLTEAHKALFYKNGN---SDFQALDKL-----LELNDESTNKEK-----483

Db 323 SVDG---LDAEFTGSKORNRALGDDVVAIEILLDVDEWAGLKEEENRRRKPDISTRG 379

QY 484 LVDDLAFAPITHPERLCKPNSQIETEVDEVRIAQLADKYTTSDGYIFDEHDIISDEGD 543

Db 380 SPFNLRIDAIVPEVPR-----SAIKARDEQVEGQTL-----FLDDQKLGADKXP 426

QY 544 AYVTPHMGSHHWTKDLSKDEKVAQYKKGILPPSPDADVK-----ANPTG 593

Db 427 KYA---GH-----VAVLQAPQGVFSGTGILRPSAANKERQRTSSGNGQSSNNG 475

QY 594 DSAATYNRVKGKRIPLVRLP 615

Db 476 NDKPKIWNFKPSDKRVFLTAIP 497

RESULT 21

D7\_DICDI

ID D7 DICDI STANDARD; PRT; 850 AA.

AC P54682;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE CAMP-inducible prespore protein D7 precursor.

GN D7.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

OX NCBI\_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AX3;

RX MEDLINE=95080502; PubMed=7988791;

RA Agarwal A., Sloger M.S., Oyama M., Blumberg D.D.;

RT "Analysis of a novel cyclic Amp inducible prespore gene in Dictyostelium discoideum: evidence for different patterns of camp regulation.";

RL Differentiation 57:151-162(1994).

CC -!- DEVELOPMENTAL STAGE: Expressed specifically in the prespore cells.

CC -!- INDUCTION: By CAMP.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U25143; AAA73514.1; --

DR DictyBase; DDB0001927; D7.

KW Sporulation; Signal.

FT SIGNAL 1 24 POTENTIAL.

FT CHAIN 25 850 CAMP-INDUCIBLE PRESPORE PROTEIN D7.

FT DOMAIN 470 475 POLY-GLN.

FT DOMAIN 555 568 POLY-ASN.

FT DOMAIN 728 738 POLY-GLN.

SQ SEQUENCE 850 AA; 95343 MW; 13BA634CCE7AA502 CRC64;

Query Match 3.4%; Score 142; DB 1; Length 850;

Best Local Similarity 16.7%; Pred. No. 3.1;

Matches 140; Conservative 122; Mismatches 296; Indels 280; Gaps 29;

QY 15 NNEVSVIDGKQATQK-TENI--TPDEVSKREGNAEQIV--IKITQGYVTHSGDHYHY 69

Db 24 SNQISDEVGKAINOKLENIEKVEDVVQFENV-ANQVIEELKVERHQ-----RQEL 73

QY 70 NGKVPYDAIISBE-----LLMKDPNYKLKDEDIYNEKGGYVIVKDGKYYVYLKDAHA 123

Db 74 LGEVHDSLRDSATNYIWGLLDKIOSYLPKDNKVKSEAEAFSSGQNNNIGSIGDSTGA 133

QY 124 DNVRTKEEN--RQKEHSHQREGTPRNDGVALARSQGRYTTDDGYIFNADIEDTG 181

Db 134 STSPQFQSLNGLSGASQSSGSGTGGTSDSK-----TTNEAIFSSKVSITDRQ 183

QY 182 DAYIVPHGDHYHYIIPKNELASASELAFAFLSGRNLNSRYRRONSNTRTNWPVS 241

Db 184 ESII-----GVAITAKDSLGAITITGLGVSSATAKVGGQITNGRAQGVITGG 231

QY 242 SNPGTIN-----TNT-----SNNSTNSQASOSNDIDSLLKOLYKLPLSQRH- 283

Db 232 DNTGTVGRGAVTTAGAVANTVGEFLGGSRGTGSSSAGTVGNVLSDSYTSIGKIASNGNS 291

QY 284 -VESDGLVDPDAQITTSRTARGVAVPHGDH-----YHPIPYX 318

Db 292 LSETIGTGTGTLAHTFAGTDSVGVT-GPHIITKTFNLIAAGKFNSDQYIDKSYGSIPO 350

QY 319 QMSLEERJARIIPURYSNHWVDSRPEQSPQPTPEPSPQPPAPNLKIDSNSSLVSO 378

Db 351 DNEEIKKRL-----QSAHQQLQEQSP-----371

QY 379 LVKRVGEGYFEEKIGISRVYFAKDLPSFTVKNLESKLSQESVSHLTAKKNVAPRQD 438

Db 372 -----AIIYQSMKSEDLKLLDDDEVIRNTLKEMQIQRENDQIQGQNGE 412

QY 439 FYDKAYNL-LTBAHKALFXNKGNSDFQALDKLRLNDENSTNEKLVDDLLAFAPITH 497

Db 413 --DKEQLIDLQNRPGLYKNQ--QDLK-----QEKRAHQQLI-----446

QY 498 PERLKPNSQIETEVDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHHW 557

Db 447 -----NYELNLQEDQEQYELLDQ-----LYDEQQ-----471

QY 558 KDSLSDEKVAQAQYTKKGILPPSPDADVKANPTGDSAAIYNRVKGEKRIPLVRLPYM 617

Db 472 ----QOPQVSNKQQLQEQIINSPED-----IQY 497

QY 618 VERTVEVKNGLIIPKDHRYHNIKFAWFDHTYKAPNGYTLDELDFATIKYYVHEPDERP 677

Db 498 LNLHNG-----PFQDDYHNDQTEELKD-----DDYFNDDQQLNNGQFENNVEEFP 543

QY 678 SNDGKNASEHVLGKDHSEDNKPKADEBPVEETPAEPVQVETEKVEAQLKAEVL 737

Db 544 LNDANDNFEQVNN 591

QY 738 LAKVTSSLSKANATETLAGLRNNLTQIMDNNSINAAEKLALLKGNPSSVSKEKI 795

Db 592 LQEVED-----APERLYEIEHNSNL-----NKAVQEAEEIERQQNGNGSPAVNSHKI 638

RESULT 22

TANA\_XENLA

ID TANA\_XENLA STANDARD; PRT; 1744 AA.

AC Q01550;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Tanabin.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Tadpole head;

RX MEDLINE=92398961; PubMed=1524825;

RA Hemmati-Brivanlou A., Mann R.W., Harland R.M.;

RT "A protein expressed in the growth cones of embryonic vertebrate neurons defines a new class of intermediate filament protein.";

```

QY 235 TNWPSVS-----NPGTNTNTSNN 254
DB 336 BEMKSIENPFSEDKAHLLQFNKEFEVRHDLFEKKLOKHFDVAKDTUNVGLRNTTVLS 395
QY 255 SNT-----NSQASQSDIDSLKQLKPLSQRH-----VESDGLV-- 290
DB 396 SNTETMLKQYEDIKENLEQKMSKSSKDEMAKTINELSVTKGLMGVQBELLTSSGNIQT 455
QY 291 -----FDPAQITRTARGVA-VPHGDHYHFIPIYSOMSELEERTARIIPLYR 336
DB 456 ALVSEMNTROELDDA---SQTKAYASLENLVRKAYAEVQSNEYERIKHLESER-- 510
QY 337 SNHWVPDRPEQSPQPTPEPSGPGAPNLKIDSNSLSVLQVRKVGEGYVFEKGLSR 396
DB 511 -----STLSSQKQIISLQ-----TK 527
QY 397 YVFADLPSETVKNLESK---LSKQESVSHLTAKENVAPRQDFYD---KAYNLLTBH 451
DB 528 EAQYEDL---VKLEAKNIEISQISGKEQSLSQTEKNENLSNELKKVQDQLEKLNLT 583
QY 452 KALPANK--GRNSFOAL-----DKLLRLNDESTNKEKLVDDLLAFAPITHPERLGP 505
DB 584 KSNYENKISSQNEIVKALVSENDTLQRIQQLVEIKENEQKD-----HTTKLEAFQKN 637
QY 506 SQIETEDEVIRIAQLADKYTSDGVIPEHDIISDEGDAYVTPHMGHSHWIGKDSL 565
DB 638 EQLQ--KLVNVEVQI-----KAHELEEQNRHLKNCLEKKTGVESLSV 683
QY 566 KVAQAQYTKKGILPPSPADVKA-----NPTGDSAAIYNVRKGEKRIPIVRLPY 616
DB 684 TL-----KQVIVLSEKQDITAEKLELQDNLESLEEVTKNLQKQVQSKR----ELEQ 733
QY 617 MVEHTEVKNGLIIPKHQHNKFNKFAWFDHDTYKAPNGYILEDLFAIKYVVEPDRP 676
DB 734 KIKELEEIKHKNRNEPSKGTQN-----FTKPSDPSKXNATNSLFPNNSAAIHSPMK 788
QY 677 HNSDGNWASBEHLVKDHSDEPNKNF-----KADBEPEVETAEPEVQVEVEKYEA 729
DB 789 PKVD-----HISKSRINSKETSKEFNDEFDLSSSNDLELTNPSP-----IQIKPVRG 837
QY 730 QLKEA-----EVLAKVTDSSLK 747
DB 838 KIKGSGNCMKPPISRSKKLLLVEDDQSLK 867
RESULT 20
ID STS5 SCHPO STANDARD; PRT; 1066 AA.
AC 074454; 013452;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein sts5.
GN STS5 OR SPCC16C4.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=97041712; PubMed=8886983;
RA Toda T., Niwa H., Nemoto T., Dhut S., Eddison M., Matsusaka T.,
RA Yanagida M., Hirata D.;
RT "The fission yeast sts5+ gene is required for maintenance of growth
RT polarity and functionally interacts with protein kinase C and an
RT osmosensing MAP-kinase pathway."
RL J. Cell Sci. 109:2331-2342(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;

```

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McComb M.W., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: Required for the maintenance of cell shape during  
 CC interphase. Required for localization of cortical actin to the  
 CC growing tips before mitosis.  
 CC -!- SUBUNIT: Interacts with serine/threonine phosphatase ppei, protein  
 CC kinase C and an osmosensing MAP kinase.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the ribonuclease II (RNB) family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D58421; BAA23619.1; -;  
 CC EMBL; AL031535; CAA20748.1; -;  
 CC PIR; T41099; T41099.  
 CC PIR; T45283; T45283.  
 CC GenDB\_Spombe; SPCC16C4.09; -;  
 CC InterPro; IPR001900; Ribonuclease\_II.  
 CC Pfam; PF00773; RNB; 1.  
 CC PROSITE; PS01175; RIBONUCLEASE\_II; FALSE\_NEG.  
 CC Hydrolase; Nuclease; Exonuclease.  
 FT CONFLICT 5 794 F -> C (IN REF. 1).  
 FT CONFLICT 794 794 V -> G (IN REF. 1).  
 SQ SEQUENCE 1066 AA; 117602 MW; 077187800B330C15 CRC64;  
 Query Match 3.4%; Score 142.5; DB 1; Length 1066;  
 Best Local Similarity 21.9%; Pred. No. 4;  
 Matches 110; Conservative 69; Mismatches 192; Indels 131; Gaps 25;  
 QY 186 VPHGDHYHFIPIKNELASLEAAAE-AFLSGR-----GNLSNRTYRQNSDNTSR 234  
 DB 55 LPLOQHMQLRHTGL-----LPVSESSFVGHRRRSASAGVGMGNFNQATI---PSNSPAV 107  
 QY 235 TNWPSV-----SNFGTTNTNTNNSNTNSQASQSDIDSLKQLKPLS----- 280  
 DB 108 SMLQPTGGGQPLVPTNFTTSVSASSDFNSPTVPS-----KFLNFSVATS 156  
 QY 281 -----QRHVESDGLVDPDAQITSTRGAVVPHGDHYHFIPIYSOMSELE--ERIAIPL 333  
 DB 157 TNISPRRHAHSHSV-----ASVSSPNSHN-AVPFTPHAFVPPVNNASPLPALNTLQQLRP 211  
 QY 334 YRSNHWVPDRPEQSPQPTPEPS-PG-----PQAPNLKIDSNSLSVQLYR 381



DR PIR: C64232; C64232.  
 DR TIGR: MG292; --; 1.  
 DR HAMAP: MF\_00036; --; 1.  
 DR InterPro: IPR002318; tRNA-synt 2c.  
 DR InterPro: IPR006193; tRNA synt Ala.  
 DR Pfam: PF01411; tRNA-synt 2c; 1.  
 DR PRINTS: PR00980; TRNASYNTHALA.  
 DR TIGRFAMS: TIGR00344; alas; 1.  
 DR PROSITE: PS50860; AA\_TRNA\_LIGASE II ALA; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 SQ SEQUENCE 900 AA; 104301 MW; AA54520BFB3949A2 CRC64;

Query Match 3.4%; Score 143.5; DB 1; Length 900;  
 Best Local Similarity 19.7%; Pred. No. 2.9; Mismatches 250; Indels 159; Gaps 28;  
 Matches 127; Conservative 109; Mismatches 250; Indels 159; Gaps 28;

QY 273 QLYKLPLSRHVESDGLVDPDPAQITRTARG-----VAVPHGDHYHPIPSQMSSELERIA 328  
 Db 199 EIWNIVFSQNDGNGNYTELAKNIDTGAIBRLVSVLQNS-----PTNFDTDIFLKL 253  
 QY 329 RII-----PURYSNHWVPSRPEQSPQPTPEP-----SPGQPPAPNL 367  
 Db 254 KIIAEAPCPKYDPSYFTF-----DPQVKYEQSYFRIIADHFKAITFTTISEGVLEGN- 306  
 QY 368 KIDNSLSVQLVR-----KVGEGYVPEEKGISRYVFAKD-----LPSETVK- 409  
 Db 307 ---ERNVVRLLRRALIAACKLQLNLAFIEKIIDEIIASVYNYOHLKAKNETVAKQVVL 363  
 QY 410 ---NLESKLSKQESVSHTLTAKENAVAPRDQEPYDKAYNLLTEAHKALFPKNGRNSDFQA 466  
 Db 364 KEINAFNKITDGLVFEKSVKNTLTPTQLNITYGFPVEIIRELVNQKGLTIDMTV 423  
 QY 467 LKLL-----LERLNDESTNKEKLVDDLLAFLAPITHPERLQKPNQSI---EYTEDVRIA 518  
 Db 424 FQQLMAKHSRSKQNNQNTINFEKQINLVNFKTKSTFFYHKNKINAKVIGLGFENLVPVK 483  
 QY 519 QLADK-----YTTSDGVIFDEHDI--SDEGDAVVT-----PHMGHSHW--I 556  
 Db 484 ELNQGSGYVFDQTVIATSGGRYDEGSCINHSNNDDKISFGQVFKPNQKHFFYFLV 543  
 QY 557 GKDSLSDK-----EKVAAQAYTK-----KGILPPSPDADVK-----ANPTGDSAAIYN 601  
 Db 544 GSKFLNDQVTLSDHETWRKLAANNHLEHLHAALQKEIDPLIKQSGAFKSAQAATIDFN 603  
 QY 602 RVKGEKRIPLVLPYVVEHTVEVKG-----NLIIPKDYH--HNI-KFAW 644  
 Db 604 LNRHLTRNELEKVENKIRSLIKQKISSKEIFTFEGSQKLNALAYPEEYSQHEILVR 663  
 QY 645 FDDHTYK-----APNGYITLEDLFAIKYVVEHPDERPHSDGWNASBHVGLGKDHSD 698  
 Db 664 FGDYSVELCGGTHVANTASIEDCFITDFYSL-----GAGWRRIEISS---NET 709  
 QY 699 PKNKFKADEPVEETPAEPE--VPQVETEKVBAQLKEAE-----VLLAKVTDSLLKANA 750  
 Db 710 INNYLKAENQKLTQLKSELEKVLSDISSIFKVELKQLQRLDKFLLPEKITQ---LRDA 766  
 QY 751 TETLAGRLNLTQLQINDNSIMAEA-----EKLALLKGSNPSSV 790  
 Db 767 SDTLALRKNDINQLTKNKYVQQQALALSITKQLLSLVDENKSYV 811

## RESULT 19

ZIPL\_YEAST  
 ID ZIPL\_YEAST STANDARD; PRT; 875 AA.  
 AC F31111;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Synaptonemal complex protein ZIPL.  
 GN ZIPL OR YDR285W OR D9819.9.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BR1824-3B;  
 RX MEDLINE=93161412; PubMed=7916652;  
 RA Sym M., Engebrecht J.A., Roeder G.S.;  
 RT "ZIPL is a synaptonemal complex protein required for meiotic  
 chromosome synapsis";  
 RL Cell 72:365-378(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Required for meiotic chromosome synapsis and cell cycle  
 progression. May act as a molecular zipper to bring homologous  
 chromosomes in close apposition. ZIPL may encode the transverse  
 filaments of the synaptonemal complex.  
 CC -!- SUBCELLULAR LOCATION: SYNAPSED MEIOTIC CHROMOSOMES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L06487; AAA35239.1; --  
 DR EMBL; U51031; AAB64474.1; --  
 DR PIR; S70115; S70115  
 DR GeneOnline; 140777; --  
 DR SGD; S0002893; ZIPL.  
 DR GO; GO:0000795; C:synaptonemal complex; IDA.  
 DR GO; GO:0007126; P:meiosis; IMP.  
 DR GO; GO:0007129; P:synapsis; IMP.  
 KW Nuclear protein; Meiosis; Coiled coil.  
 FT DOMAIN 177 333 COILED COIL (POTENTIAL).  
 FT DOMAIN 397 438 COILED COIL (POTENTIAL).  
 FT DOMAIN 456 752 COILED COIL (POTENTIAL).  
 FT CONFLICT 55 55 T -> A (IN REF. 1).  
 SQ SEQUENCE 875 AA; 100035 MW; 674F12625CD9DDED CRC64;

Query Match 3.4%; Score 142.5; DB 1; Length 875;  
 Best Local Similarity 18.9%; Pred. No. 3;  
 Matches 176; Conservative 117; Mismatches 304; Indels 333; Gaps 41;

QY 3 ELGLQART-VKENRVSYIDGKQATQKTENLTPDESKREGINAEQIVIKITDQGYVTS 61  
 Db 86 EIGSPKTTSTDOYNRL-----KNDVALENDTDEDPEITEVREYSEGVAKETKE----S 137  
 QY 62 HGD-----HYHYNGKVPYDAII-----SEELLM----- 85  
 Db 138 HGDPNDSSETTLKDSKRMHEYTMGTNGKAPLHTSINNGSTSSNDVLLFAFTQTORICSNLQOE 197  
 QY 86 ----KDPNKKK-----DEDIVNEVGKGVIVKDGKYYVYVKD-----AAHADN 125  
 Db 198 LQOQODNAKLKVRLOSVAASDKINE-----KV-GKYSCLTLOERATLTSHKNN 249  
 QY 126 VRTKEEINRQKQESHQSHREGGTPRNDGAVALARSQGRYTTDDGYTFNADSIIDETG---- 181  
 Db 250 QETKLKDLQNHQLYQRRISG-----FKTS---TENLKTINDLGKNNK 290  
 QY 182 ---DAYIVPHGDHYYIPKNELSAELAAAEAFSLGRGNLSNRTYRRQNSD-----NTSR 234  
 Db 291 EADAELMKKGKIEYL-KRELDCCS-----GQLSEKI---KNSSLIQEMGKNR 335

KW	Alternative splicing; Phosphorylation; Coiled coil.	
FT	DOMAIN 58 177	
FT	COILED COIL (POTENTIAL).	
FT	DOMAIN 437 555	
FT	COILED COIL (POTENTIAL).	
FT	MOD RES 266 266	
FT	PHOSPHORYLATION (BY MAPK).	
FT	MOD RES 276 276	
FT	PHOSPHORYLATION (BY MAPK).	
FT	MOD RES 287 287	
FT	PHOSPHORYLATION (BY MAPK).	
FT	VARSPPLIC 201 201	
FT	Missing (in isoform 1a and isoform 1d).	
FT	/FtId=Vsp.002775.	
FT	S -> SPRQSWRKS (in isoform 1b and isoform 1e).	
FT	VARSPPLIC 201 201	
FT	/FtId=VSP.002776.	
FT	Missing (in isoform 1a and isoform 1b).	
FT	VARSPPLIC 219 249	
FT	/FtId=Vsp.002777.	
FT	Missing (in isoform 3a).	
FT	VARSPPLIC 410 415	
FT	/FtId=Vsp.002778.	
FT	Missing (in isoform 3a).	
FT	VARSPPLIC 505 513	
FT	/FtId=VSP.002779.	
FT	R -> G: RESULTS IN INHIBITION OF JNK BINDING.	
FT	MUTAGEN 205 205	
FT	P -> G: RESULTS IN INHIBITION OF JNK BINDING.	
FT	MUTAGEN 206 206	
FT	T -> G: RESULTS IN INHIBITION OF JNK BINDING.	
FT	MUTAGEN 207 207	
FT	S -> G: RESULTS IN INHIBITION OF JNK BINDING.	
FT	MUTAGEN 208 208	
FT	L -> G: RESULTS IN INHIBITION OF JNK BINDING.	
FT	MUTAGEN 209 209	
FT	T -> A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPKBP3; WHEN ASSOCIATED WITH A-276 AND 287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY.	
FT	MUTAGEN 266 266	
FT	T -> A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPKBP3; WHEN ASSOCIATED WITH A-266 AND A-287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY.	
FT	MUTAGEN 276 276	
FT	T -> A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPKBP3; WHEN ASSOCIATED WITH A-266 AND A-287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY.	
FT	MUTAGEN 287 287	
FT	K -> R (IN REF. 4).	
FT	CONFLICT 312 312	
FT	E -> L (IN REF. 3; AAF26843).	
FT	CONFLICT 376 376	
FT	F -> K (IN REF. 4).	
FT	CONFLICT 561 561	
Query Match	3.5%; Score 144; DB 1; Length 1337;	
Best Local Similarity	19.4%; Pred. No. 4.5; Mismatches 112; Indels 271; Gaps 40;	
Matches 164;		
Qy	81 BELMKDPNYK-LKDEDIVNEKGGYIKVDGKYVY-----LKDAHADNV-RTKEEI 132	
Db	97 EKALRKQAEKFIPEFEDALEQEKELQIWE-HYEFQTRQLELKAKNYAQISRLERE 154	
Qy	133 NRQKE-----HSQRE-----GGTPNDGAVALARSQGRYTTDDGYIF 171	
Db	155 SEMKEYNALQRHTEMIQTYVEHIERSKMQVGGQGTSSLP-GRSKERPTSLNVFP 213	
Qy	172 NASDIETGDAYIVPHGDHYIIPKNLSASELAEEAFISGRLNSRTYRQNSDN 231	
Db	214 LADGWRAQMGKLVVPAGDWH-----LSDLGQSSSSYQCPN-DE 254	
Qy	232 TSRTNWVPSVNPQTNTNTSNNTNSQASQNSDIDSLKQLYK----- 277	
Db	255 MSEGQSSAAATPTTGT-----KSNTPTSVPSAAVTELSLOPLGDSVSVTKNKQAR 310	
Qy	278 -PLSQRVESGLVDFPAQITSRARGVAVPHGDHYHPIFYQSOMSELEERARIPIRYR 336	
Db	311 EKRSRNME-----VQVTT-QEMRNVSIGWSS-----DEWSDVQDIDISTPELDV- 354	
Qy	337 SNHWVPSDRPQPSQPPT-----EPPSGP----- 361	
Db	355 ----CPETRLERTSSPTQGVNKAFGINTDSLHYELSTAGSEVIGDVDEGADLLGPVS 410	
Qy	362 -----QPAPNLKIDNSSLV-----SOLVRKV-----GEGYVFEKG-ISRIV 398	

Db	411 RDPFGMGKEVGNLLLE-NSQLETKNALNVVKNDLIAKVDPQLSGEQEVL--KGELEAK 467	
Qy	399 FAKDLPSETVKNLESKSKQESVSHTLTAKENVAPRQDEYDKAYNLLTAHAKALFNK 458	
Db	468 QAKVLENRIKELEELKRVK--SEAVTARRE---PR--EEVEDVSSYLCTELDKIPMAQR 521	
Qy	459 GRNSDFQALDKLLERLNDESTNKEKLVDDLLAFLAPITHPERLKGPKNSQIEYTEDEVRIA 518	
Db	522 RRFTREVMARVLMER-----NQYK-----ERLMELQEAQWTE-MIRAS 559	
Qy	519 QLADKYTTSDGYIFDEHDIISDEGDAYVTPMNGSHHWIGKDSLSDEKKEKVAQAQYTK----- 574	
Db	560 R-----EHPVSQEKKSTIWFQSRLEF-----SSSSSPPPAKRSPVSNH 600	
Qy	575 -----EKGILPPS--PDADVKNAPTGDSSAAAIYRNVKGEKRIPLV 612	
Db	601 YKSPTAAGFSQRRSHALCQISAGSRPLEFFPDDDCTSSARREQKREQTRQVR----- 652	
Qy	613 RLPYVVEHTVEVKNGLIIPHKDHYHNIKFAWFDHDTYK--APNGYTTLEDL---PATIKY 667	
Db	653 -----EH---VRN-----DDGRLQACGMSLPKYLQSLPNG-GQEDTRMKNVPVPV 694	
Qy	668 YVEHPDERPHNSDNGWGNASEHVLGKHSDPDKNFK-----ADEEPVETTP 714	
Db	695 YCRPLVEKDPSTKLWCAAGVNLGSKWPKHEEDSSNGPKPVPGRDPLTCDREGEGEPEKSTHP 754	
Qy	715 AEPEVPQVETKEVAQLKEAEVLLAKYVTDSS--LKANATETLAGLRNRLTLQIMNNSI 771	
Db	755 SPEKKAKETPEADATSSRVWLITSTLTSTKVIIDANQPGTIIVDOFTVCNAHVLICISSI 814	
Qy	772 MAEAE 776	
Db	815 PAASD 819	
RESULT 18		
SYA_MYCGE	SYA_MYCGE STANDARD; PRT; 900 AA.	
ID	P47534;	
AC	01-FEB-1996 (Rel. 33, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--trna ligase) (Alars).	
GN	ALAS OR MG292.	
OS	Mycoplasma genitalium.	
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.	
OX	NCBI_TaxID=2097;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ATCC 33530 / G-37;	
RX	MEDLINE=96026346; PubMed=7569993;	
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,	
RA	Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,	
RA	Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,	
RA	Nguyen D.T., Uitterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,	
RA	Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Lucier T.S.,	
RA	Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,	
RT	"The minimal gene complement of Mycoplasma genitalium.";	
RL	Science 270:397-403(1995).	
CC	-!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +	
CC	diphosphate + L-alanyl-tRNA(Ala).	
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.	
CC	-!- SIMILARITY: Belongs to Class-II aminoacyl-tRNA synthetase family.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
DR	EMBL; U39709; AAC71513.1; -.	

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1A), FUNCTION, PHOSPHORYLATION, AND  
 RP INTERACTION WITH MAPK8; MAPK9; MAPK10; MAP2K4 AND MAP3K1.  
 RC TISSUE=Brain;  
 RX MEDLINE=99455010; PubMed=10523642;  
 RA Ito M., Yoshioka K., Akechi M., Yamashita S., Takamatsu N.,  
 RA Sugiyama K., Hibi M., Nakabeppu Y., Shiba T., Yamamoto K.-I.;  
 RA "JSAPl, a novel jun N-terminal protein kinase (JNK)-binding protein  
 RT that functions as a scaffold factor in the JNK signaling pathway";  
 RL Mol. Cell. Biol. 19:7539-7548 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1A; 1B; 1C; 1D AND 1E), AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Brain;  
 RX MEDLINE=20480689; PubMed=11024282;  
 RA Ito M., Akechi M., Hirose R., Ichimura M., Takamatsu N., Xu P.,  
 RA Nakabeppu Y., Tadayoshi S., Yamamoto K.-I., Yoshioka K.;  
 RT "Isoforms of JSAP1 scaffold protein generated through alternative  
 RT splicing";  
 RL Gene 255:229-234 (2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1C AND 3A), FUNCTION, INDUCTION, TISSUE  
 RP SPECIFICITY, SUBCELLULAR LOCATION, PHOSPHORYLATION, MUTAGENESIS OF  
 RP ARG-205; PRO-206; THR-207; SER-208; LEU-209; THR-266; THR-276 AND  
 RP THR-287, AND INTERACTION WITH MAPK8IP2; MAPK8; MAPK9; MAPK10; MAP2K7  
 RP AND MAP3K11.  
 RC STRAIN=C57BL/6; TISSUE=Brain, and Heart;  
 RX MEDLINE=20094982; PubMed=10629060;  
 RA Kelkar N., Gupta S., Dickens M., Davis R.J.;  
 RA "Interaction of a mitogen-activated protein kinase signaling module  
 RT with the neuronal protein JIP3";  
 RL Mol. Cell. Biol. 20:1030-1043 (2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1C), SUBCELLULAR LOCATION, AND INTERACTION  
 RP WITH KLC1.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=20560743; PubMed=1106729;  
 RA Bowman A.B., Kamal A., Ritchings B.W., Philp A.V., McGrail M.,  
 RA Gindhart J.G., Goldstein L.S.B.;  
 RA "Kinesin-dependent axonal transport is mediated by the Sunday Driver  
 RT (SYD) protein";  
 RL Cell 103:583-594 (2000).  
 RN [5]  
 RP SEQUENCE OF 1240-1337 FROM N.A.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.-I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettum M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,  
 RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [6]  
 RP INTERACTION WITH KLC.  
 RC TISSUE=Brain;  
 RX MEDLINE=21135887; PubMed=11238452;  
 RA Verhey K.J., Meyer D., Deehan R., Blenis J., Schnapp B.J.,

RA Rapoport T.A., Margolis B.;  
 RT "Cargo of kinesin identified as JIP scaffolding proteins and  
 RT associated signaling molecules";  
 RL J. Cell Biol. 152:959-970 (2001).  
 CC -1- FUNCTION: The JNK-interacting protein (JIP) group of scaffold  
 CC proteins selectively mediates JNK signaling by aggregating  
 CC specific components of the MAPK cascade to form a functional JNK  
 CC signaling module. May function as a regulator of vesicle  
 CC transport, through interactions with the JNK-signaling components  
 CC and motor proteins.  
 CC -1- SUBUNIT: Forms homo- or heterooligomeric complexes. The central  
 CC region of Mapk8ip3 interacts with the C-terminal of Mapk8ip2 but  
 CC not Mapk8ip1. Binds specific components of the JNK signaling  
 CC pathway namely Mapk8, Mapk9 and Mapk10 to the N-terminal region,  
 CC Map2k4 and Map2k7 to the central region and Map3k11 to the C-  
 CC terminal region. Binds the TPR motif-containing C-terminal of  
 CC kinesin light chain, pre-assembled Mapk8ip1 scaffolding complexes  
 CC are then transported as a cargo of kinesin, to the required  
 CC subcellular location.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; localised in the soma and  
 CC growth cones of differentiated neurites and the Golgi and vesicles  
 CC of the early secretory compartment of epithelial cells.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Name=1c; Synonyms=3b;  
 CC IsoId=Q9ESN9-1; Sequence=Displayed;  
 CC Name=1a;  
 CC IsoId=Q9ESN9-2; Sequence=VSP\_002775, VSP\_002777;  
 CC Name=1b;  
 CC IsoId=Q9ESN9-3; Sequence=VSP\_002776, VSP\_002777;  
 CC Name=1d;  
 CC IsoId=Q9ESN9-4; Sequence=VSP\_002775;  
 CC Name=3a;  
 CC IsoId=Q9ESN9-5; Sequence=VSP\_002778, VSP\_002779;  
 CC Name=1e;  
 CC IsoId=Q9ESN9-6; Sequence=VSP\_002776;  
 CC -1- TISSUE SPECIFICITY: Highly expressed throughout many regions of  
 CC the brain and at lower levels in the heart, liver, lung, testes  
 CC and kidney. All isoforms have been identified in the brain,  
 CC Mapk8ip3A is also expressed in the spleen and lung.  
 CC -1- INDUCTION: Expressed in neurites 5 days following initiation of  
 CC nerve growth factor Ngf induced differentiation. Ngf withdrawal  
 CC results in the down-regulation of Mapk8ip3 protein by caspase-  
 CC mediated cleavage.  
 CC -1- SIMILARITY: Belongs to the JIP scaffold family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB005662; BAB58974.1; -  
 CC EMBL; AB043124; BAB16675.1; -  
 CC EMBL; AB043125; BAB16676.1; -  
 CC EMBL; AB043123; BAB16674.1; -  
 CC EMBL; AB043129; BAB16685.1; -  
 CC EMBL; AB043126; BAB16685.1; JOINED.  
 CC EMBL; AB043127; BAB16685.1; JOINED.  
 CC EMBL; AB043128; BAB16685.1; JOINED.  
 CC EMBL; AF178637; AAF26843.1; -  
 CC EMBL; AF178636; AAF26842.1; -  
 CC EMBL; AF262046; AAG36931.1; ALT\_INIT.  
 CC EMBL; BC004003; AAH04003.1; -  
 CC MGD; MGI:1353598; Mapk8ip3.  
 CC GO; GO:0005737; Cytoplasm; IEP.  
 CC GO; GO:0019894; F:kinesin binding; IPI.  
 CC GO; GO:0005078; F:MAP-kinase scaffold activity; IPI.  
 CC GO; GO:0019901; F:protein kinase binding; IPI.  
 CC GO; GO:0046328; P:regulation of JNK cascade; IDA.  
 CC GO; GO:0016192; P:vesicle-mediated transport; IDA.

